이학박사학위논문

Efficient and specific genome editing in human cells using CRISPR/Cas system

2014 년 2 월

서울대학교 대학원

화학부 생화학 전공

조 승 우

Abstract

Efficient and specific genome editing in human cells using CRISPR/Cas system

Seung Woo Cho

Department of Chemistry

The Graduate School

Seoul National University

Programmable nucleases based on the FokI nuclease domain have been used widely for genome targeting in various cells and organisms. Recent identification of a prokaryotic immune system called CRISPR shows that a

CRISPR/Cas system has the prospect of being another programmable nuclease for genome targeting in other higher eukaryotes.

In this study, I have used a CRISPR/Cas system as a programmable RNA-guided endonucleases (RGENs), which cleave DNA in a targeted manner for genome engineering in human cells. These enzymes can induce site specific double strand breaks (DSBs) in chromosomal DNA and the repair of DSBs gives rise to targeted genome modifications via non-homologous end joining or homologous recombination. In this study, RGENs can induce efficient mutations at frequencies of up to 75% in human K562 cells. Compared to ZFNs or TALENs, the specificity of RGENs can easily be customized by synthesizing new guide-RNA molecules.

Furthermore, I investigated off-target effects of RGENs. Off-target cleavage at unwanted loci can cause genetic instability including tumorigenesis or chromosome structural variations. The present study demonstrates that RGENs can efficiently discriminate on-target sites from off-target sites that differ by two bases. Indeed, exome sequencing analysis shows that RGEN-induced mutant clone without off-target mutations can be isolated. These features will make RGENs the most powerful and applicable tool for genome engineering.

Keywords: Genome Editing, Engineered Nucleases, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), CRISPR-Accociated protein (Cas), RNA-guided Endonucleases (RGENs)

Student Number: 2007-20326

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Cas	CRISPR associated protein
CRISPR	clustered regularly interspaced palindromic repeats
DSB	double strand break
HR	homologous recombination
IDLV	integrase deficient lenti virus
NHEJ	non-homologous end joining
NLS	nuclear localization signal
PAM	protospacer adjacent motif
RGEN	RNA-guided endonuclease
TALEN	transcription activator like effecter nuclease
ZFN	zinc finger nuclease

I. Introduction

Genome manipulation is an essential technique for studying genome and diseases. Although gene targeting by spontaneous homologous recombination (HR) has been used to alter genome sequence for several decades (CAPECCHI 1989), it is a time- and effort-consuming process due to its low efficiency (SEDIVY and SHARP 1989). There are different methods for manipulating genome with specific enzyme such as integrase (MILLER and ROSMAN 1989) or transposase (GEURTS *et al.* 2003; RUBIN and SPRADLING 1982); however, it is based on random integration (SCHRODER *et al.* 2002; WU *et al.* 2003) which can cause cancer or other defects (HACEIN-BEY-ABINA *et al.* 2003).

Genome targeting efficiency could be greatly enhanced by the DSB repair process (ROUET *et al.* 1994). It is well known that the yeast homing endonuclease, I-SceI, can promote HR with high efficiency (>1,000 fold) at targeting locus (JASIN 1996); however, there is no endogenous target sequence of rare-cutting restriction enzymes in human or other higher species.

The first programmable restriction enzyme, zinc finger nuclease (ZFN), consists of the zinc finger domain which is well-characterized. It is one of the most common DNA binding motifs and the non-specific DNA cleavage domain of type IIs restriction enzyme, FokI (KIM *et al.* 1996). DNA binding

specificity of ZFN is determined by its zinc finger composition and it can be customized by fusing three to six fingers (KIM and PABO 1998) of several hundreds of naturally occurring (BAE *et al.* 2003) or engineered zinc finger domains (REBAR and PABO 1994).

Carroll showed ZFN could induce a DSB at the desired locus to manipulate the genome of *Drosophila melanogaster* with highly enhanced efficiency (BIBIKOVA *et al.* 2003). As a result of this milestone study, ZFN have been used to induce site-specific DSBs to modify the genomes in a targeted manner in cells (HOCKEMEYER *et al.* 2009; KIM *et al.* 2009; URNOV *et al.* 2005) and organisms (CUI *et al.* 2011; DOYON *et al.* 2008; GEURTS *et al.* 2009; MENG *et al.* 2008; MORTON *et al.* 2006; SHUKLA *et al.* 2009; TOWNSEND *et al.* 2009; YANG *et al.* 2011; YU *et al.* 2011).

However, broad applications of ZFNs were hampered by the lack of a convenient and rapid method for the design of functional ZFNs (MAEDER *et al.* 2008). Indeed, some of the ZFNs induced cytotoxicity by off-target cleavage (URNOV *et al.* 2010). Transcription activator like effecter nucleases (TALENs) showed improved efficiency and specificity than ZFN (KIM *et al.* 2013; MILLER *et al.* 2011); however, a variety of materials and sub-cloning steps were needed to synthesize TALENs (REYON *et al.* 2012; SANJANA *et al.* 2012).

The clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein (Cas) system has a role for adaptive immunity to viruses and plasmids in bacteria and archaea (WIEDENHEFT *et al.*

2012). The ribonucleoprotein (RNP) complex of type II CRISPR/Cas system in *Streptococcus pyogenes* wihch consists of CRISPR RNA (crRNA), trans-activating RNA (tracrRNA) and Cas9 protein, performs as a sequence-specific endonuclease that cleaves foreign genetic material to protect the host cells. In *S.pyogenes*, pre-crRNAs are transcribed at CRISPR locus containing spacers and direct repeat sequences, and are further processed into mature crRNAs by RNase III (DELTCHEVA *et al.* 2011). Recent papers have shown that the crRNA did function as a guide-RNA whose specificity is determined by spacer sequences (JINEK *et al.* 2012). Indeed, the specificity of crRNA is programmable by replacing the 20 base sequences of 5' upstream of them. The specificity of an RGEN is determined by the 20 base pair sequences in crRNA and the NGG trinucleotide, known as the protospacer adjacent motif (PAM) (MOJICA *et al.* 2009), recognized by Cas9 itself. The 22 base pairs of specificity might be sufficient to ensure the specific genome targeting in higher eukaryotes.

Here, I present the evidence that RGENs can induce site-specific genome modifications in human cells at high frequencies (CHO *et al.* 2013a). RGEN did not induce severe cytotoxicity and significant off-target effects (CHO *et al.* 2013b). These results propose that RGEN can be a versatile and a convenient tool for genome manipulation in cells and organisms.

II. Materials and Methods

1. Construction of Cas9 expression plasmid

The Cas9 coding sequence (4,104 bp), derived from *Streptococcus pyogenes* strain M1 GAS (NC_002737.1), was reconstituted using the human codon usage table. Humanized Cas9 sequence was converted to oligonucleotide to synthesis (ROUILLARD *et al.* 2004). First, 1-kbp DNA segments were assembled using overlapping 35-mer oligonucleotides and Phusion polymerase (New England Biolabs), and cloned in to T-vector (Solgent). A full length Cas9 sequence was assembled using four 1-kbp DNA segments by overlap PCR and sequenced by Sanger sequencing (Figure 1).

The Cas9 open-reading frame was subcloned into p3s, which was derived from pcDNA3.1 (Invitrogen). In this vector, Cas9 protein is expressed under the control of the CMV promoter. To ensure protein expression and compartmentalization, HA-epitope and a nuclear localization signal (NLS) were attached at the N-terminus or C-terminus of Cas9 protein (Figure 2).

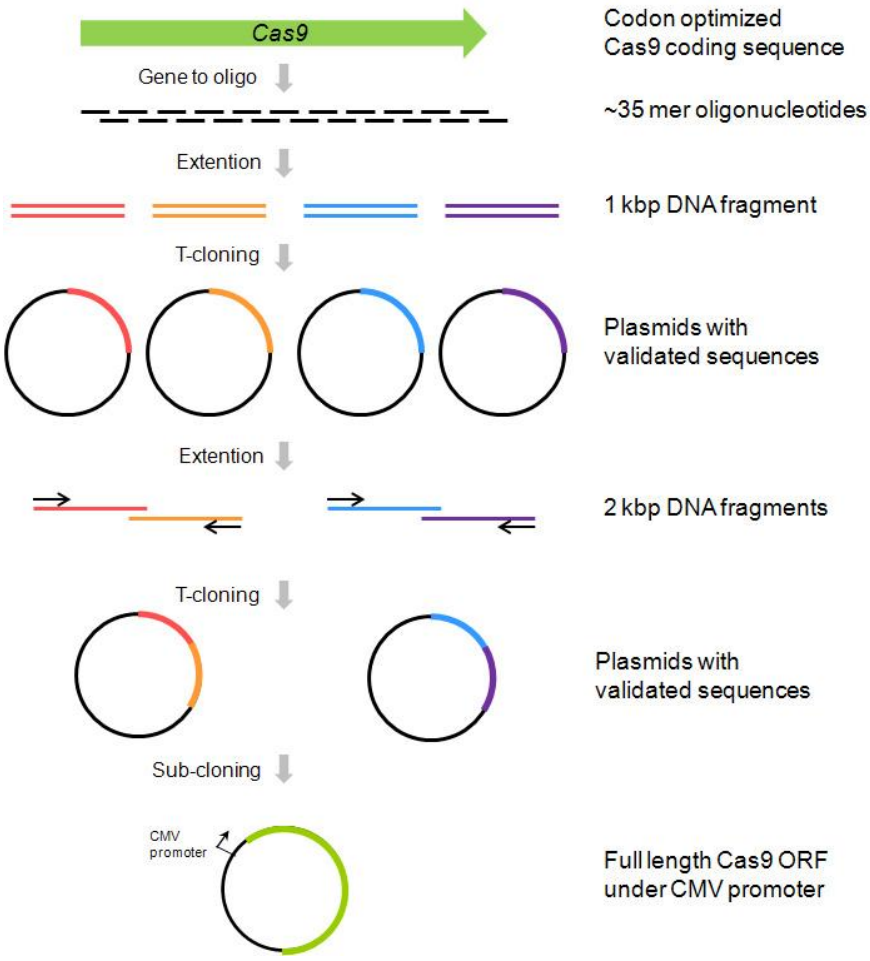


Figure 1. Schematic representation for construction of the Cas9-encoding plasmid. The Cas9 coding sequence was confirmed by Sanger sequencing.

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPE
DNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ
SITGLYETRIDLSQLGGD<u>GGSGPPKKKRKVYPYDVPDYA</u>*

Figure 2. Amino-acid sequence of the expressed version of Cas9 used in this study. The amino-acid sequence of the Cas9 protein encoded in our Cas9 expression plasmid is presented. The sequence of the NLS (PPKKKRKV) and the HA-epitope (YPYDVPDYA) at the C-terminus is underlined.

2. Construction of guide-RNA transcription plasmid

Human U6 promoter was used to drive expression of RNA components of CRISPR/Cas9 system because of its defined transcription start and termination site (LOBO *et al.* 1990). U6 promoter was sub-cloned from pLKO.1-puro plasmid into pUC18 using AatII/EcoRI restriction sites. For convenient replacement of target sequence, type IIs restriction sites were inserted into backbone plasmid, pRG2-CT (Figure 3)

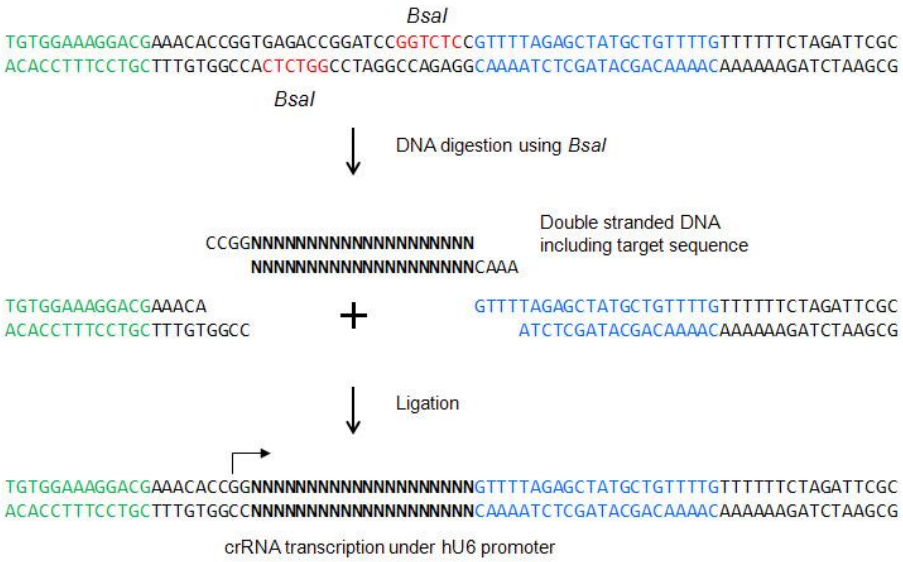


Figure 3. DNA construct for crRNA transcription under hU6 promoter. BsaI type IIs restriction enzyme sites (Red) were employed to generate sequence-independent overhang. The 3' end of hU6 promoter and backbone sequence of crRNA are shown in green and blue, respectively. The arrow indicates the known transcription start site of hU6 promoter.

3. Cas9 protein purification

The Cas9 expression cassette was sub-cloned into pET28-b(+) and transformed into BL21(DE3) strain. The expression of Cas9 was induced using 0.2 mM IPTG for 16 hr at 18℃. The Cas9 protein containing the His_6-tag was purified using Ni-NTA agarose beads (Qiagen) and diazyed against 20 mM HEPES pH 7.5, 150 mM KCl, 1 mM DTT, and 10% glycerol (JINEK *et al.* 2012). The Purified Cas9 protein was concentrated using Ultracel 100K cellulose column (Millipore). The purity and concentration of Cas9 were analyzed by electrophoresis using 8% SDS-PAGE gel.

4. Preparation of guide-RNA

RNA was in vitro transcribed through run-off reaction by T7 polymerase using MEGAshortscript T7 kit (Ambion) according to the manufacturer's manual. Templates for RNA in vitro transcription were generated by annealing two complementary oligonucleotides. To synthesize large amount of guide-RNA, 200 nM of template DNAs were incubated with 10 U/μl of T7 RNA polymerase, 0.4 U/ μl of yeast inorganic pyrophophatase, 4 mM of each NTP and 10 U/μl of RNase inhibitor in a reaction buffer including 40 mM Tris-HCl pH 7.9, 20 mM $MgCl_2$, 2mM spermidine and 10 mM DTT. RNA was purified and recovered in nuclease-free water followed by phenol:chloroform extraction, chloroform extraction, and ethanol precipitation.

Purified RNA was quantified by spectrometry and analyzed by electrophoresis using 8 % denaturing urea-PAGE gel.

5. In vitro cleavage assay

170-1,700 ng of purified Cas9 protein was incubated with 350 ng of plasmid DNA or 200 ng of PCR amplicon with 50-500 ng of RNAs in a reaction volume of 20 μl in NEB buffer 3 for 1 hrs at 37℃. Digested DNA was visualized and analyzed by electrophoresis with agarose gels.

6. Cell culture and transfection

K562 (ATCC, CCL-243) cells were grown in RPMI-1640 with 10% FBS and a penicillin/streptomycin mix (100 U/ml and 100 μg/ml, respectively). 2×10^6 K562 cells were transfected with 20 μg of Cas9-encoding plasmid using the 4D-Nucleofactor, SF Cell Line 4D-Nucleofactor X kit, and Program FF-120 (Lonza) according to the manufacturer's instruments. After 24 hrs, 60 μg of crRNA and 120 μg of tracrRNA were transfected into 1×10^6 K562 cells.

HEK293/17 cells (ATCC, CRL-11268) or HeLa cells (ATCC, CCL-2) were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 100 units/ml penicillin, 100 μg/ml streptomycin, 0.1 mM nonessential amino acids, and 10% FBS. 0.8×10^5 HeLa cells were transfected with 500 ng of Cas9-encoding plasmid and 500 ng of tracrRNA/crRNA expression plasmid using Lipofectamine 2000 (Invitrogen) according to the

manufacturer's protocol. Cells were collected 2-9 days after transfection and genomic DNA was isolated.

7. Preparation of nuclear extract and western blotting

2×10^6 K562 cells transfected with Cas9-encoding plasmids were harvested after 48 hrs transfection. 50 ul of Buffer A (10 mM HEPES-KOH pH 7.9, 10mM KCl, 0.1mM EDTA, 1mM DTT, 0.4% IGEPAL, and protease inhibitor cocktail) was added to cell pellet. After disrupting cell pellet by pipetting, the mixture was adapted to centrifugation at 16,000 g for 3 min. The supernatant was saved as a cytosolic fraction. The residual pellet was washed with Buffer A twice, then 50 ul of Buffer B (20mM HEPES-NaOH pH 7.9, 400 mM NaCl, 1mM EDTA, 10% glycelol, 1mM DTT and protease inhibitor cocktail) was added to pellet. The re-suspended pellet was incubated for 2 hrs at 4 ℃. After centrifugation, supernatant was saved as a nucleus fraction. The concentration of each fraction was measured by Bradford assay. Proteins were resolved on 8% SDS-PAGE gel and adapted to western blotting. All antibodies used in western blotting were purchased from Santa Cruz.

8. Validation of mutations

Purified genomic DNAs from RGEN-treated human cells were adapted to analysis for RGEN-induced mutagenesis. The region including the targets or potential off-target sites were amplified using primers (Table) and Taq

polymerase. The amplicons were denatured and re-annealing to from heteroduplex using thermocycler and a following program (95℃ for 2 min, -2℃/s to 85℃, -0.1℃/s to 25℃). 10 μl of re-annealed DNA was incubated with 5U of T7E1 mismatch specific enzyme and NEB buffer 2 in a reaction volume of 20ul for 20 min at 37℃, and then analyzed by electrophoresis using 1.5% agarose gels. The mutation frequency was calculated based on the fraction of cleaved DNA which was measured by densitometry (KIM *et al.* 2009).

9. Immunostaining

HeLa cells transfected with the *C4BPB*-specific RGEN were fixed on glass slides and incubated first with anti-TP53BP1 rabbit polyclonal antiboby (Bethyl Laboratories) and then with Alexa Fluor 488-conjugated secondary antibody (Invitrogen). Cells were mounted in the presence of DAPI (sigma) and examined under a fluorescence microscope (Zeiss).

10. Targeted deep sequencing and analysis

Genomic regions harboring on-target and potential off-target sites of RGENs were amplified using Phusion polymerase and subjected to paired-end read sequencing using Illumina MiSeq at Bio Medical Laboraories. Insertion or deletion located around the RGEN cleavage site was considered as a signature of RGEN-mediated mutagenesis.

11. Exome sequencing and analysis

Genomic DNAs from RGEN-induced mutant clones were used to construct exome-captured libraries using a TrueSeq Exome Enrichment kit (Illumina). PCR enriched libraries were subjected to Illumina HiSeq 2000 at Bio Medical Laboratories. Raw Reads in FASTQ format were aligned to the NCBI36/Hg18 reference genome using BWA-0.5.9. PCR duplicates were removed using SAMtools-0.1.16. Single nucleotide variants and indels were called using GATK-1.4. Since engineered nucleases rarely induce substitution via DSB repair process (KIM *et al.* 2009), only indel variants were considered in further analysis as previously described. Shared variants found in at least two non-relative clones were excluded. Further, naturally-occurring common indels were also excluded by comparing to dbSNP build 130. Indel variants presented in wild-type K562 cells were not considered as RGEN-induced mutations.

Remained indel variants were analyzed by two methods. First, Amplicons harboring the indel were subjected to in vitro cleavage assay. Second, sequences around indel junction were compared to the on-target sequences to find putative off-target sites.

Table 1. Primer sequences used in T7E1 assay.

Gene	Direction	Sequence (5' to 3')
CCR5	F1	CTCCATGGTGCTATAGAGCA
	F2	GAGCCAAGCTCTCCATCTAGT
	R	GCCCTGTCAAGAGTTGACAC
C4BPB	F1	TATTTGGCTGGTTGAAAGGG
	R1	AAAGTCATGAAATAAACACACCCA
	F2	CTGCATTGATATGGTAGTACCATG
	R2	GCTGTTCATTGCAATGGAATG
ADCY5	F1	GCTCCCACCTTAGTGCTCTG
	R1	GGTGGCAGGAACCTGTATGT
	F2	GTCATTGGCCAGAGATGTGGA
	R2	GTCCCATGACAGGCGTGTAT
KCNJ6	F	GCCTGGCCAAGTTTCAGTTA
	R1	TGGAGCCATTGGTTTGCATC
	R2	CCAGAACTAAGCCGTTTCTGAC
CNTNAP2	F1	ATCACCGACAACCAGTTTCC
	F2	TGCAGTGCAGACTCTTTCCA
	R	AAGGACACAGGGCAACTGAA
N/A Chr. 5	F1	TGTGGAACGAGTGGTGACAG
	R1	GCTGGATTAGGAGGCAGGATTC
	F2	GTGCTGAGAACGCTTCATAGAG
	R2	GGACCAAACCACATTCTTCTCAC

Table 2. Oligonucleotide sequences used in vitro transcription.

For short chimeric guide-RNA transcription

Direction	Sequence (5' to 3')
F	GAAATTAATACGACTCACTATAGG [Target sequence 20 mer] GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG
R	CGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC [Target complementary sequence 20 mer] CCTATAGTGAGTCGTATTAATTTC

For crRNA transcription

Direction	Sequence (5' to 3')
F	GAAATTAATACGACTCACTATAGG [Target sequence 20 mer] GTTTTAGAGCTATGCTGTTTTG
R	CAAAACAGCATAGCTCTAAAAC [Target complementary sequence 20 mer] CCTATAGTGAGTCGTATTAATTTC

For tracrRNA transcription

Direction	Sequence (5' to 3')
F	GAAATTAATACGACTCACTATAGGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCG
R	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATG

Table 3. Primer sequences used in Sanger sequencing.

Name	Sequence (5' to 3')
CMV-proF	AAATGGGCGGTAGGCGTG
F632	AGCTGTTCATCCAGCTGGTGCAGACCTACAACCA
F1316	CCCACCAGATCCACCTGGGCGAGCTGCAC
F1959	GTTCGAGGACCGCGAGATGATCGAGGAGCGC
F2827	GGCTTCATCAAGCGCCAGCTGGTGGAGACCC
F3473	AGTACGGCGGCTTCGACAGCCCCACCGT
F4168	CTGAGCCAGCTGGGCGGCGACGGCGGCT
Bla-R	GGGAATAAGGGCGACACGGAAATG

III. Results

1. Establishment of RGENs for genome targeting

a. In vitro cleavage assay

In prokaryotes, crRNA (42 mer) and tracrRNA (75 mer) are generated from the host genome (Figure 4) (DELTCHEVA *et al.* 2011). A single-chain chimeric RNA (62 mer) produced by fusing an essential portion of crRNA and tracrRNA can replace the two RNAs in the Cas9/RNA complex to form a functional endonuclease (JINEK *et al.* 2012). In order to test the DNA cleavage activity of Cas9 in human cells, the target sequence was selected in human *CCR5* gene, which encodes an essential receptor of HIV and is a target for the therapy of AIDS (PEREZ *et al.* 2008).

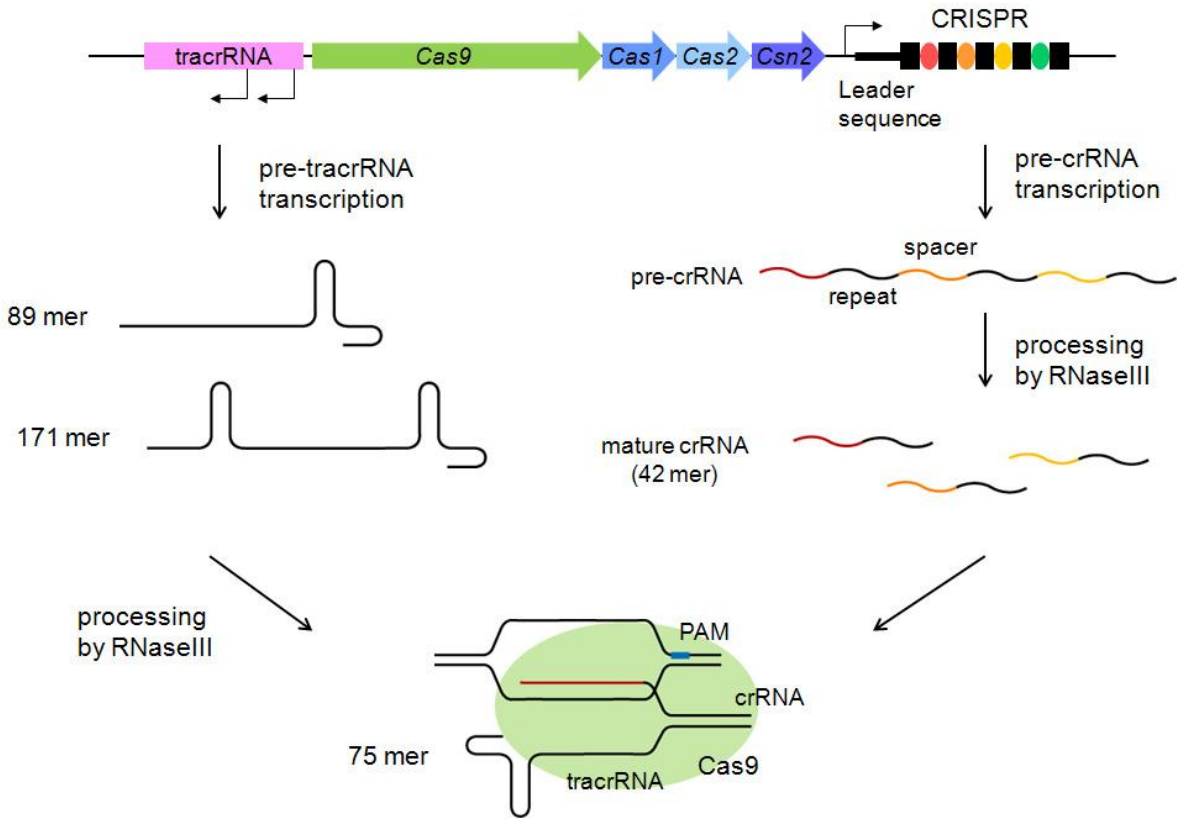


Figure 4. Schematic diagram for CRISPR biogenesis in *S.pyogenes*.

A Cas9 target sequence consists of a 20 bp DNA sequence complementary to the crRNA and the trinucleotide (5'-NGG-3') PAM recognized by Cas9 itself. Purified recombinant Cas9 was incubated with an in vitro transcribed chimeric-RNA and a plasmid harboring the 23 bp target sequence. Cas9 efficiently cleaved the plasmid at the expected position only in the presence of the synthetic guide-RNA and target sequence (Figure 5).

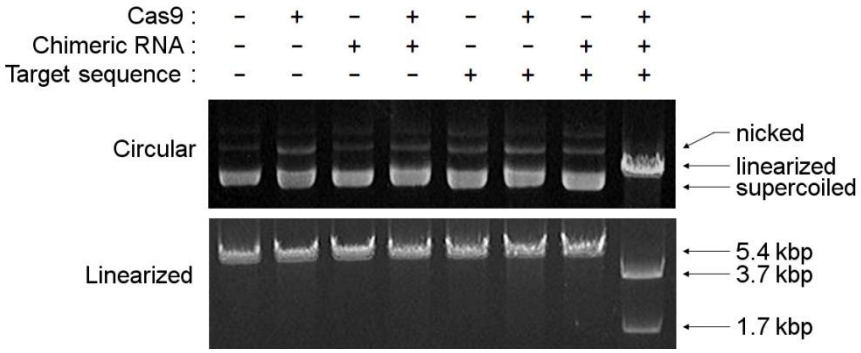


Figure 5. In vitro cleavage of plasmid DNA by Cas9. An intact circular plasmid or pre-linearized plasmid was incubated with Caas9 and chimeric RNA.

b. RGEN-mediated mutagenesis in human cells

For efficient cleavage of chromosomal DNA, Cas9 has to be compartmentalized to nuclear. To demonstrate this, protein extract of each fraction of cell was isolated and subjected to western blotting. As expected, the presence of Cas9 in both nuclear and cytosol was confirmed by anti-HA-probe (Figure 6).

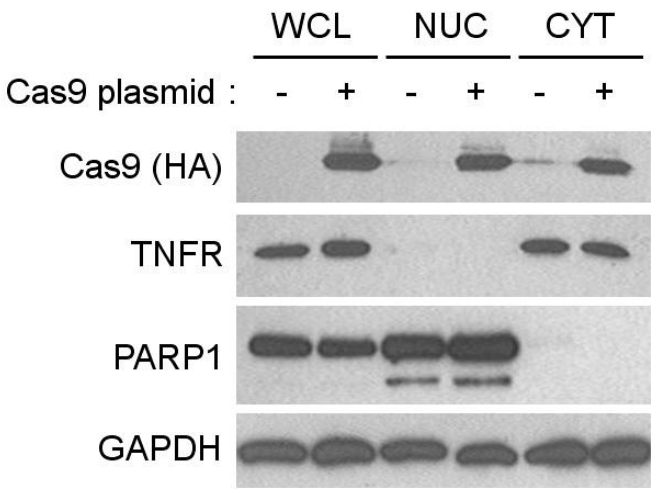
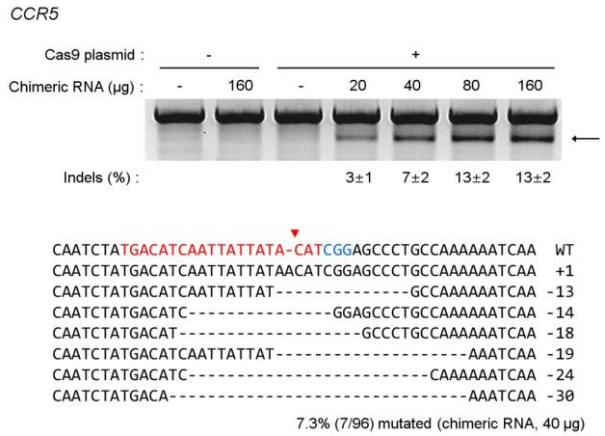


Figure 6. Expression and compartmentalization of the Cas9 protein in human cells. Protein expression in each fraction of K562 cells was determined by western blot. Antibodies against HA-epitope, TNFR, PARP1, or GAPDH were used as makers for Cas9, cytosol extract (CYT), nuclear extract (NUC), and whole cell lysate (WCL), respectively.

To test whether RGENs could be used for targeted disruption of endogenous genes in human cells, human K562 cells were transfected with Cas9-encoding plasmid and a synthetic chimeric-RNA targeting *CCR5*. Genomic DNA was isolated from transfected cells and subjected to T7E1 assay. Cleaved bands, indicate that the induced mutations, were found only when the cells were co-transfected with both Cas9 and RNA (Figure 7). Mutation frequencies, estimated from the relative DNA band intensities, were RNA-dosage dependent. DNA sequencing analysis of PCR amplicons corroborated the induction of RGEN-mediated mutations at the targeted site.

Next, *CCR5* targeting guide-RNA was replaced with a newly synthesized guide-RNA designed to target the human *C4BPB*, which encodes the beta chain of C4b-binding protein, a transcription factor. As shown in *CCR5*, this RGEN induced mutations at the chromosomal target site in K562 cells. Sequencing analysis of PCR amplicons revealed that, out of four mutant sequences, one clone contained a single-base insertion, and one a two-base insertion precisely at the expected cleavage site. These results indicate that RGENs can cleave chromosomal DNA at the expected position as well as in vitro cleavage assay.

A



B

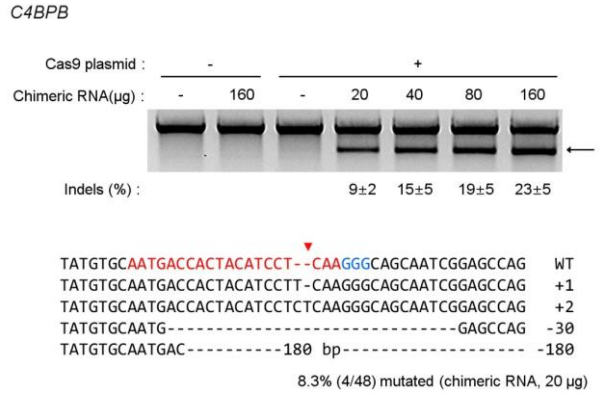


Figure 7. RGEN-driven mutations at endogenous chromosomal sites. (A) *CCR5* and (B) *C4BPB* locus. The mean frequencies and standard error of the mean of independent experiments are shown. The regions of the target sequence complementary to the guide-RNA and the PAM sequence are shown in red and blue, respectively. The red triangle indicates expected cleavage site. The column of the right indicates the number of inserted or deleted bases.

c. Guide-RNA structure

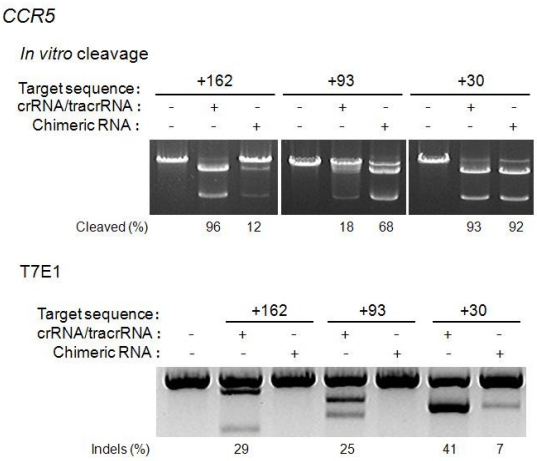
In order to investigate if RGEN could be generally used to target human genome, an additional 18 genes of human genome were targeted. However, there were no RGENs that could induce targeted mutations up to the detection limits (~1%). I speculated that the single-chain chimeric guide-RNAs were active as well as natural forms (dual guide-RNA) of crRNA and tracrRNA in vitro (Figure 8). However, RNA structure or stability may be affected by the outside sequences of essential region.

Figure 8. Schematic diagram of short chimeric guide-RNA and dual guide-RNA. Sequences derived from crRNA and tracrRNA were shown in green and red, repectively.

To compare the activity of chimeric RNAs and dual guide-RNAs, three in vitro transcribed crRNAs and tracrRNAs or chimeric RNAs were employed in an in vitro assay (Figure 9). One of the *CCR5* targeting RGEN showed no difference at cleavage efficiency. Another RGENs showed enhanced or reduced efficiency when the natural forms of RNAs were used.

To test whether different RNA structures affect DNA targeting efficiency in cells, four RGENs were introduced into K562 cells. When the chimeric RNAs were tested, only one out of three targets showed mutations. Interestingly, dual guide-RNAs showed drastically enhanced efficiencies at all target sequences examined (Figure 9). These results suggest that essential regions were not sufficient for efficient gene targeting.

A



B

Figure 9. Comparison of short chimeric guide-RNA and dual guide-RNA. (A) Two types of guide-RNA against the three target sequences in *CCR5* were compared in vitro (Upper) and in cellular (Bottom). (B) Identification of mutatation sequences induced by crRNA/tracrRNA/Cas9 complex. Each number of target sequences means the distance from start codon in coding sequence.

2. Analysis of off-target effects of RGENs

a. Analysis of RGEN-induced cytotoxicity

Non-specific engineered nucleases showed several numbers of foci against DSB repair regulator by immunostaining (MILLER *et al.* 2007). In order to test the specificity of RGEN, the number of DSBs in an RGEN-transfected cell was analyzed using anti-TP53BP1, a regulator of DSB repair response (O'DRISCOLL and JEGGO 2006). In contrast to positive control, RGEN-treated cells showed only ~1 DSB per cells (Figure 10). Thus, RGEN did not induce non-specific DSBs frequently.

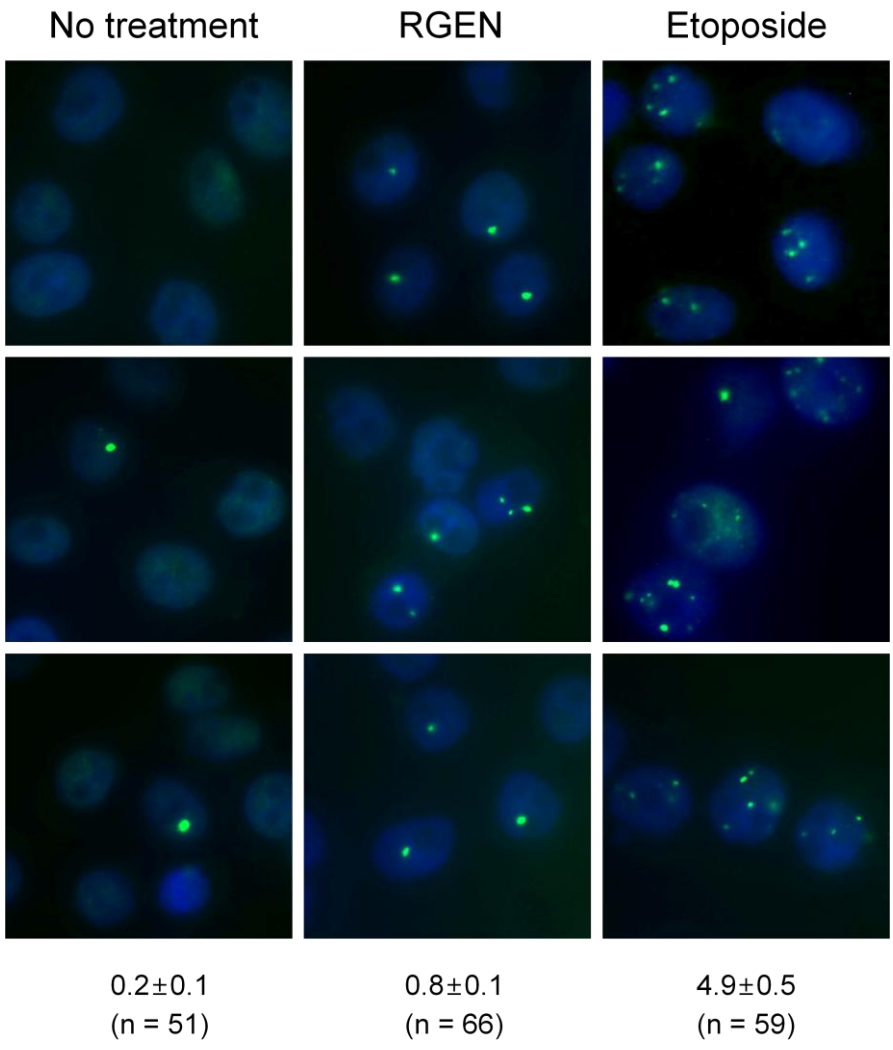


Figure 10. RGEN-induced DSBs in cells detected by TP53BP1 staining. HeLa cells transfected with the *C4BPB*-specific RGEN were fixed and adapted to immunostaining by anti-TP53BP1 antibody. DAPI (blue) and TP53BP1 (green) images were merged. Etoposide (1 uM) was used as a positive control. The average number of foci per cell and the standard error of the mean are shown at the bottom of each picture.

If engineered nucleases induce non-specific DSBs in a cell, it causes cytotoxic effect which give rises to cell death (PRUETT-MILLER *et al.* 2008). Thus, mutations become reduced with the passage of time. To investigate whether RGEN could induce a cytotoxic effect in a cell, mutation frequencies were measured with the course of time. Mutations induced by RGEN were stably maintained, even at 8 days post-transfection (Figure 11). These results indicate that active RGEN did not cause cytotoxicity in human cells.

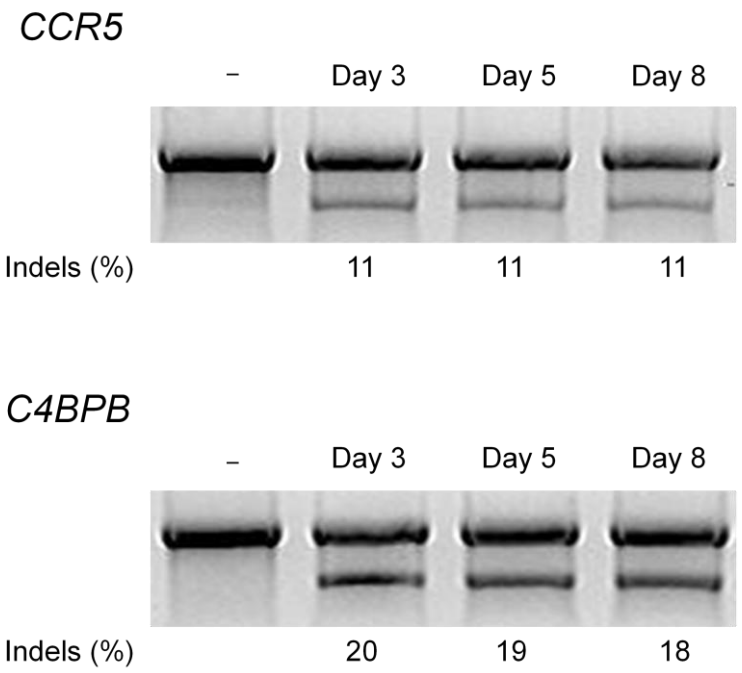


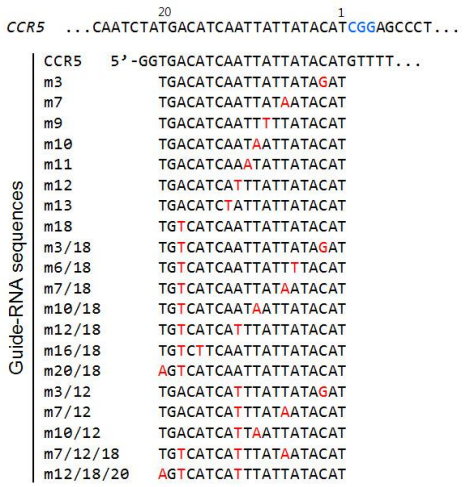
Figure 11. Stable maintenance of RGEN-induced mutant cells. Mutation frequencies were measured using the T7E1 assay at 3, 5, and 8 days post-transfection.

b. Analysis of RGEN specificity using mismatched guide-RNA

Although RGEN showed no obvious cytotoxicity, recent studies showed that Cas9 complexed with a guide-RNA that contains a one-base mismatch to the target sequence can cleave the target DNA both in human (CONG *et al.* 2013), and in bacterial cells (JIANG *et al.* 2013), and in vitro (JINEK *et al.* 2012). RGENs tolerate mismatches especially in the 5' upstream region, but not in the seed region of 6- to 11- nucleotides that is immediately upstream of the PAM sequence. In order to test whether *CCR5* targeting RGEN tolerate the mismatches, crRNAs with one to three mismatches were synthesized and introduced into K562 cells (Figure 12A).

It was observed that mismatches in 8 bp of 5' upstream were tolerable (Figure 12B). This was consistent with the results of previous reports. However, when one additional mismatch was introduced within 11 bp upstream of the PAM, RGEN mediated mutation frequencies was lower than the limit of detection. Surprisingly, three mismatches were not tolerable, even though they were all located in 8 bp of 5' upstream.

A



B

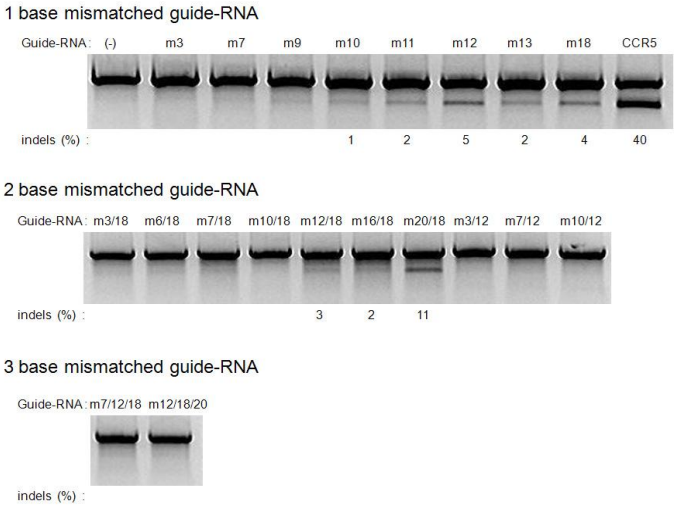
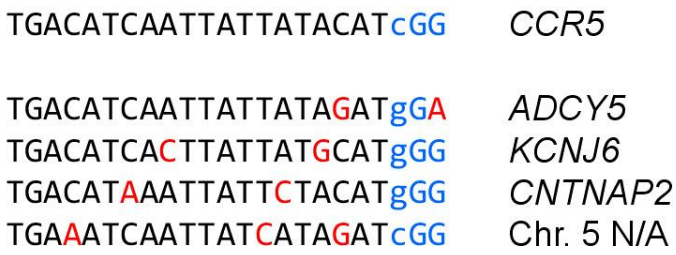


Figure 12. Specificity of RGEN using mismatched guide-RNA. (A) Sequence of endogenous *CCR5* target (Top) and the mismatched crRNA variants (Bottom). (B) T7E1 assay for K562 cells transfected with Cas9 plasmid and the guide-RNAs listed in (A)

c. Analysis of off-target effects of RGENs at homologous sites

The experiments using mismatched guide-RNAs were convenient for exploring the specificity of RGEN, however, it has a critical limitation that the mismatched RNAs were not verified for activities against to their on-target sequences. Thus, to investigate the specificity of RGEN, mismatched target sequences have to be examined using identical guide-RNAs. To this end, the most homologous sequences to target sequences of *CCR5* were searched for in the human genome (Figure 13A). T7E1 assays showed that mutations were not induced at these sites (Figure 13B): three sites with two mismatches each, and one site with three mismatches.

A



B

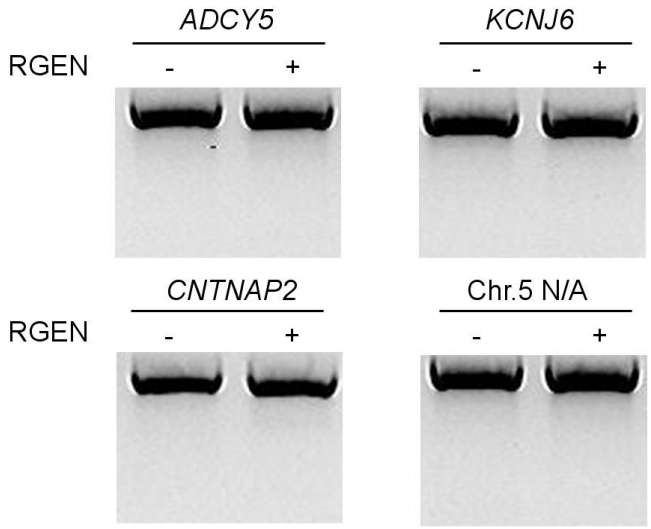


Figure 13. Undetectable off-target mutations at homologous sites of *CCR5* targeting RGEN. (A) Homologous sequence of *CCR5* on-target sequence. Mismatched sequences were shown in red and the PAM sequences were shown in blue. N/A, not applicable, an intergenic site. (B) T7E1 assay at those sites.

To examine the off-target effect of RGEN more systematically, potential off-target sequences were searched with the following categories: 1) sites with conserved sequences in seed region; or 2) sites with one to three mismatches in the entire 22 bases. Due to poor sensitivity of T7E1 assay, further analysis was addressed by targeted deep sequencing. The two RGENs induced indels at frequencies of 75% and 60% at corresponding on-target sites of *C4BPB* and *CCR5* in K562 cells, respectively. In sharp contrast, no indels were detected at any measurable frequency at any of the putative off-target sites (Figure 14). *CCR5* targeting RGEN did not induce any mutations at two sites (off15 and off16) that each contained a one-base mismatch in the 11-bp seed region and another one-base mismatch in the 5' upstream region. Indeed no mutations were found even at Off1 sites of *C4BPB*, whose three mismatches are all located in the 5' upstream region. Note that the off-17 site contained a single-base mismatch in the 20-bp sequence corresponding to the crRNA and another single-base in the PAM. No indels were detected at this site, suggesting that PAM recognition by Cas9 is also critical.

A B

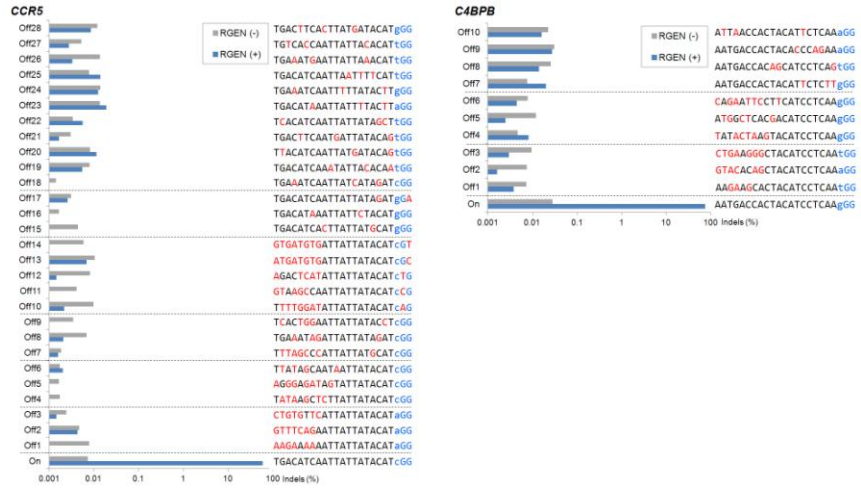
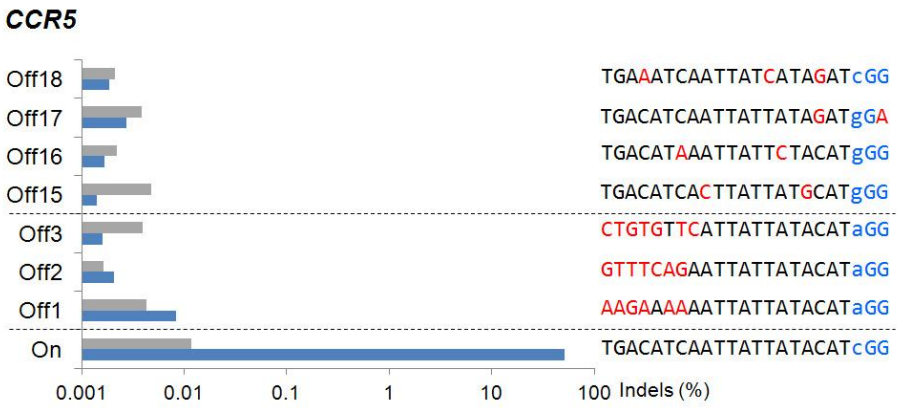


Figure 14. Mutation frequencies at on-target and potential off-target sites of the *CCR5*- (A) and *C4BPB*- (B) specific RGENs in K562 cells. Cells were transfected with crRNA, tracrRNA, and the Cas9 plasmid or the Cas9 plasmid alone (negative control). PCR amplicons that span the on-target site and potential off-target sites were subjected to deep sequencing. Sequences that contained indels around the expected cleavage site were considered to be RGEN-induced mutations.

The off-target effects were also addressed by RGEN-encoding plasmids in another human cell line, HeLa, and it was found that these RGENs discriminated on-target sites from their most likely off-target sites by a factor of >10,000-fold (Figure 15). These results suggest that the 5' upstream region outside the seed region in the target sequence also contributes to the specificity of RGEN, and that RGENs cannot efficiently clave chromosomal DNA with mismatches of two or more nucleotides, at least one of which occurs in the seed region.

A



B

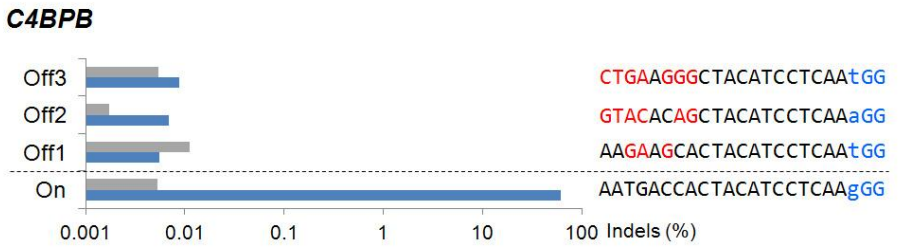
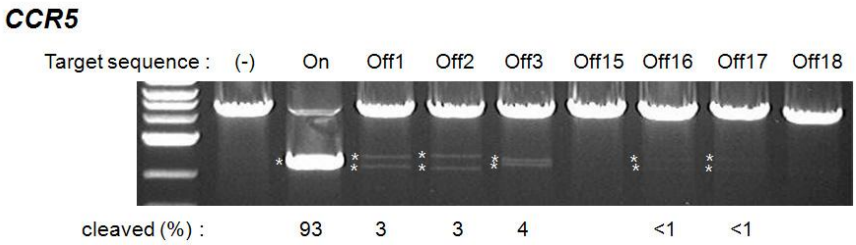


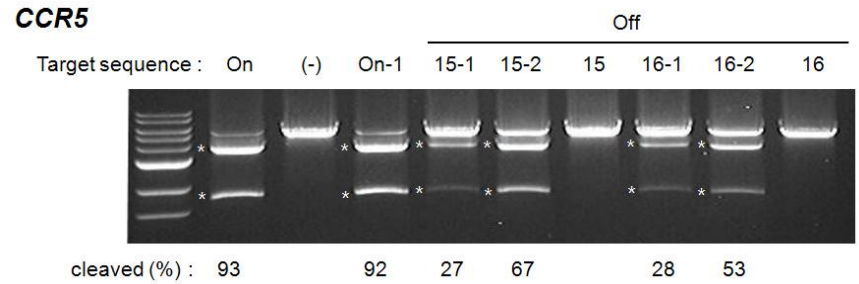
Figure 15. Mutation frequencies at on-target and potential off-target sites of the *CCR5*- (A) and *C4BPB*- (B) specific RGENs in HeLa cells.

Next, the specificity of RGENs was addressed by in vitro cleavage assay (Figure 16). Cas9/crRNA/tracrRNA complex was incubated with plasmids that contained either the *CCR5* on-target sequence or off-target sequences. Interestingly, the *CCR5*-specific RGEN showed a low level (<1%) of cleavage activity toward plasmids that contained the off-target sequences, suggesting that RGENs function less discriminatingly in vitro than they do in cells under this experimental conditions. To investigate how one- or two-base mismatches were different in manner of specificity of RGEN, it was generated that plasmids which contained artificial sequences with each single base mismatch (off15-1, 15-2, 16-1 and 16-2). These hybrid sequences were cleaved much more efficiently by the ribonucleoproein complex than were the off-target sequences with two-base mismatches (Off15 and off16). This result indicates that RGENs can distinguish on-target sites from off-target sites with two-base mismatches though not those with a single-base mismatch.

A



B



C

On	TGACATCAATTATTATACATcGG	On-1	TGACATCAATTATTATACATgGG
Off1	AAGAAAAAATTATTATACATaGG	Off15-1	TGACATCAATTATTATGCATgGG
Off2	GTTTCAGAATTATTATACATaGG	Off15-2	TGACATCACTTATTATACATgGG
Off3	CTGTGTTCATTATTATACATaGG	Off16-1	TGACATCAATTATTCTACATgGG
Off15	TGACATCACTTATTATGCATgGG	Off16-2	TGACATAAATTATTATACATgGG
Off16	TGACATAAATTATTCTACATgGG		
Off17	TGACATCAATTATTATAGATgGA		
Off18	TGAAATCAATTATCATAGATcGG		

Figure 16. In vitro cleavage assay of on-target or potential off-target sequences by the *CCR5*-specific RGEN. Plasmids that contain putative off-target (A) or hybrid (B) sequences were digested with the recombinant Cas9 protein complexed with crRNA and tracrRNA. (C) DNA sequences of the on-target, off-target and hybrid sequences.

d. Extended analysis of off-target effects of RGENs

To address the generality of specificity of RGENs, target sequences which have another endogenous site with one or two mismatches were searched in human genome (Table 4). *CCR5* was chosen for targeting because it has the most homologous gene, CCR2. An additional 9 RGENs were designed to target *CCR5* differing by one- or two- base mismatches with CCR2. The mutation frequencies were measured at on-target in *CCR5* and off-target in CCR2. Three RGENs (Figure 17A, B, and C) did not induce any off-target mutations at the CCR2 sites that carried two-base mismatches. One RGEN (Figure 17D) induced off-target mutations at the CCR2 site that carried a one-base mismatch in the seed region and another one-base mismatch in the upstream region at a frequency of 1.8%, 29-fold lower that the on-target mutation frequency.

However, RGENs cannot efficiently discriminate on-target sites from off-target sites that differ by a single base. Two RGENs (Figure 17F and G) showed undistinguishable mutation frequencies at off-target compare to on-target; two RGENs were still able to discriminate on-target sites from off-target sites with a single base mismatch by a factor of 7.2- to 42-fold. Thus, all four RGENs (Figure 17E to H) induced indels at frequencies that ranged from 1.6% to 43% at off-target sites that carried a single-base mismatch either in the seed region or the upstream region. One RGEN (Figure 17I) had an off-target site with a single base mismatch in the PAM. This RGEN discriminated

the on-target site from the off-target sites by 490-fold, again confirming the importance of PAM recognition by Cas9.

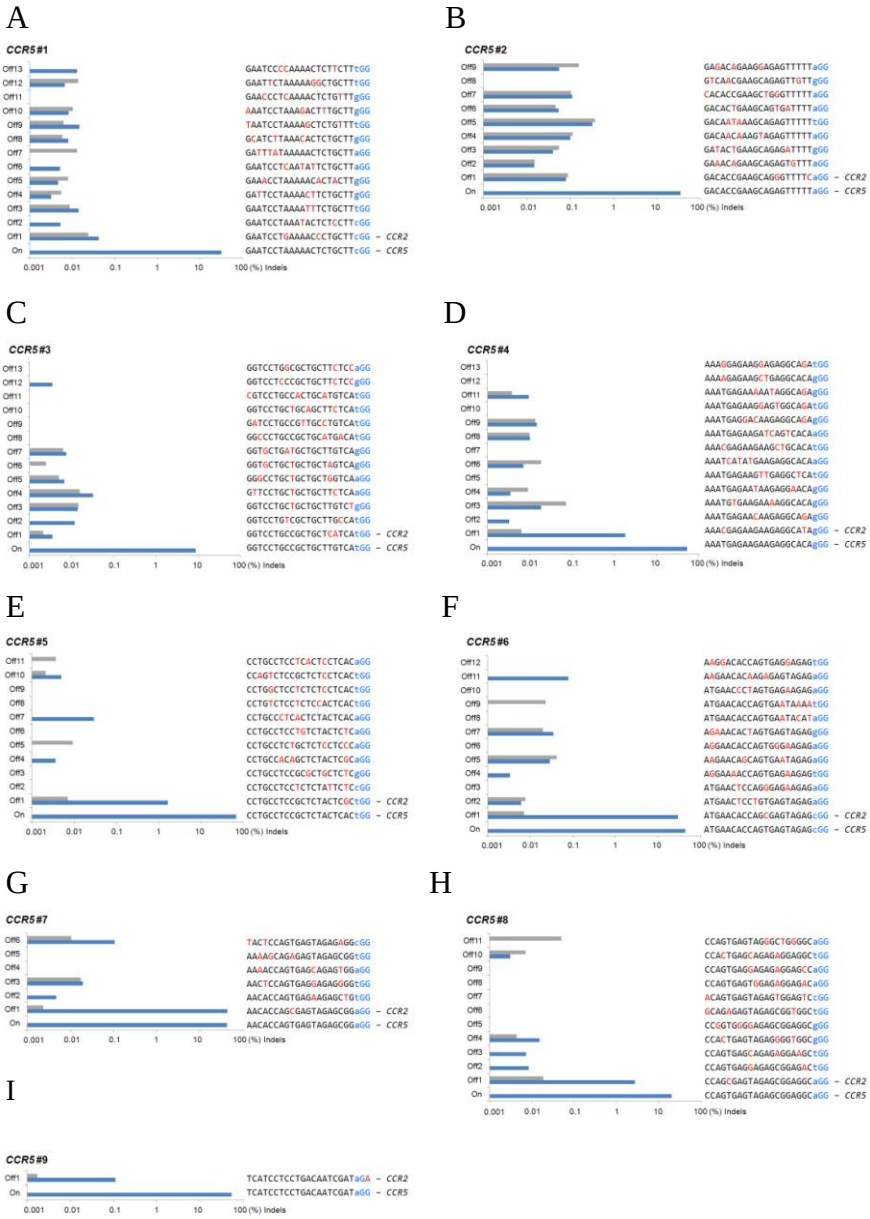


Figure 17. Analysis of RGEN-induced mutations at on-target and off-target sites. (A) - (I) K562 cells were transfected with RGENs that target nine different sites. Mutation frequencies at each locus were estimated by deep sequencing.

These RGENs did not induce any indels at several potential off-target sites identified in other loci (Figure 17). Taken together, these results suggest that RGENs can distinguish on-target sites from off-target sites that differ by at least two bases, and that one could avoid or minimize off-target effects of RGENs by choosing unique target sites that do not have homologous sequences elsewhere in the genome.

Table 4. Summary of potential off-target sites analyzed in this study. The number of in sillico-searched potential off-target sites in the human genome is shown in brackets.

	C4BPB	***CCR5***	***CCR5* #1**	***CCR5* #2**	***CCR5* #3**	***CCR5* #4**	***CCR5* #5**	***CCR5* #6**	***CCR5* #7**	***CCR5* #8**
mis0	**1(1)**	**1(1)**	**1(1)**	**1(1)**	**1(1)**	**1(1)**	**1(1)**	**1(1)**	**1(1)**	**1(1)**
mis1	**(0)**	**(0)**	**(0)**	**(0)**	**(0)**	**(0)**	**1(1)**	**1(1)**	**1(1)**	**1(1)**
mis2	**(0)**	**2(2)**	**3(3)**	**1(1)**	**3(3)**	**4(6)**	**(0)**	**1(1)**	**1(2)**	**1(1)**
mis3	**5(7)**	**11(16)**	**10(16)**	**8(9)**	**10(14)**	**9(72)**	**10(21)**	**10(15)**	**4(9)**	**9(13)**

e. Analysis of clonal exome sequencing

To examine whether the off-target mutations were induced at any unpredictable sites in human genome, exome sequencing were performed to identify off-target mutations on a genomic scale. Using limiting dilution, clonal populations were obtained which have RGEN-induced mutations in *CCR5* or *C4BPB*. Of 37 or 47 clones, 29 (78%) or 35 (74%) had mutations in the *CCR5* or *C4BPB*, respectively. These high mutation frequencies were comparable to those observed with deep sequencing. Genomic DNA isolated from one *C4BPB*-disrupted clone, two *CCR5*-disrupted clones, and wild-type K562 cells was subjected to exome sequencing. Analysis of the exome sequence data revealed a total of 308 indels present in the four clones, including wild-type cells, compared to the reference genome; 255 in *CCR5*-disrupted clone #1, 246 in *CCR5* clone #2, 233 in the *C4BPB* clone, and 253 in the wild-type clone (Figure 18) (Note that the numbers are comparable in these clones). Among these, 190 indels were present in all 4 clones including wild-type cells and 74 indels were shared in at least 2 clones. These indels were excluded from further analysis. 10 indels were found only in the wild-type clone, which suggests that the coverage of exome sequencing was not 100%. Likewise, 8, 9, and 17 indels were found only in the *C4BPB* clone, *CCR5* clone #1, and *CCR5* clone #2, respectively.

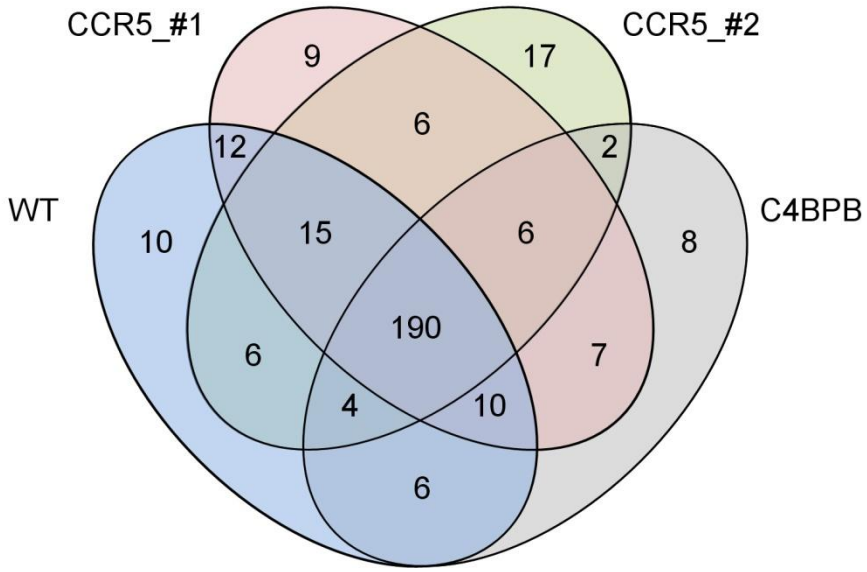


Figure 18. Validation of indels found in exome sequencing. A Venn diagram showing the number of indels found in mutant and wild-type cells.

The 13 indels reported in a public database (dbSNP build 130, NCBI) were excluded in further analysis (Table 5) (OUSTEROUT *et al.* 2013), because they are unlikely to result from RGEN-mediated DNA cleavages. Sanger and deep sequencing were used to show that most of these "clone-specific" indels are present even in wild-type cells. Then, putative off-target sequences around the remaining indels were searched in the reference genome (Table 7). Other than the on-target mutations induced by the RGENs in these clones, none of these indels were associated with potential off-target sites. It is most likely that the indels in gene-disrupted clones were the results of incomplete coverage of exome sequencing (Table 6) (CLARK *et al.* 2011). Exons are captured before sequencing but this process covers only <90% of the human exome. >10% of exons missed in wild-type cells can be captured in mutant clones.

Table 5. Summary of exome sequencing results. Indels found in mutant clones are listed. None of these indels were induced by RGEN.

Clone	Chr.#	Position	Gene	Common indel[a]	Methods of invalidation[b]	Reference Sequence	Altered Sequence
CCR5_#1 only	2	128199143	*WDR33*	-	Sanger	TCT	-
	4	1378622	*CRIPAK*	-	N/A	-	CACA
	5	5262273	*ADAMTS16*	-	Sanger	-	A
	8	37674782	*ZNF703*	-	N/A	A	-
	11	17309054	*NUCB2*	rs3842269	MiSeq	ACA	-
	12	130004994	*GPR133*	-	*In vitro*	TACTCCTGGCT GCTGC	-
	16	29820303	*ASPHD1*	-	Sanger	-	GGT
	20	226688	*ZCCHC3*	rs5839847	N/A	CGG	-
	21	9964796	*TPTE*	-	N/A	CTT	-
CCR5_#2 only	1	1876952	*KIAA1751*	-	MiSeq	G	-
	1	2928268	*ACTRT2*	rs4013154	MiSeq	AGA	-
	1	12862133	*PRAMEF4*	-	N/A	-	G
	1	26481430	*UBXN11*	-	*In vitro*	-	GGGACA
	1	77796933	*AK5*	-	MiSeq	-	TAT
	3	46389429	*CCR5*	-	Sanger	ACATCAATTATTAT	-
	3	75868732	*ZNF717*	rs67410759	MiSeq	G	-
	6	32659933	*HLA-DRB1*	rs67798196	-	G	-
	6	32659936	*HLA-DRB1*	-	Sanger	-	TT
	7	130891570	*PODXL*	rs11277659	N/A	GGCGAC	-
	11	1607775	*KRTAP5-5*	-	*In vitro*	-	GGCTGTGGCTCC
	12	110521137	*ATXN2*	rs10560189	N/A	GCT	-
	15	83420116	*PDE8A*	-	MiSeq	CTCTC	-
	17	70350725	*GRIN2C*	-	N/A	-	GCTCCGGGG
	17	76834096	*SLC38A10*	rs10569617	Sanger	ATG	-
	19	60016486	*KIR2DL4*	rs11371265	N/A	-	A
	22	19109973	*SCARF2*	rs5844418	-	-	G
Shared in CCR5_#1 and CCR5_#2	3	75869005	*ZNF717*	-	*In vitro*	CTACATTCT	-
	3	120616609	*ARHGAP31*	-	MiSeq	CACACCTGGGG	-
	3	196995726	*MUC4*	-	N/A	-	GCTGGTGACAGG AAGAGGGGTGGT GTGACCTGAGGA TGCTGAGGAAGG
	6	32659926	*HLA-DRB1*	rs67476479	N/A	G	-
	10	21845472	*SKIDA1*	-	MiSeq	-	CCTCCT
	19	39535594	*KIAA0355*	-	MiSeq	CCCCACCCCAG C	-
C4BPB only	4	47100179	*GABRB1*	-	Sanger	TTTCAATTG	-
	5	156412136	*HAVCR1*	rs45623443	Sanger	TTGGAACAGTCG TCA	-
	6	32657377	*HLA-DRB1*	rs67687262	MiSeq	C	-
	6	32657380	*HLA-DRB1*	-	*In vitro*	-	T
	11	36553378	*RAG1*	-	MiSeq	A	-
	12	49032254	*FAM186A*	-	N/A	AGAGGGATCCC CAATTCCTGAGC CTGCGGAGGGA TC	-
	19	62775469	*ZNF416*	-	*In vitro*	-	A
	X	32271298	*DMD*	-	MiSeq	-	T

[*] RsNumbers of common indels in dbSNP build 130 are listed.

[**] Sanger sequencing or deep sequencing (MiSeq) showed that these indels were present in wild-type K562 cells. In vitro, PCR products spanning these indels were tested via in vitro cleavage assay (Figure X.). N/A, not applicable.

Table 6. The mean coverage of exome sequencing in this study.

Clone	Target Base Count	Mapped Target Depth	Target Average Depth	X5 Target Region Coverage (%)
*CCR5*_#1	62,257,316	3,242,297,629	52.1	91.6
*CCR5*_#2		3,571,954,659	57.4	90.8
C4BPB		2,288,223,860	36.8	89.7
WT		2,963,014,865	47.6	91.1

To rule out the possibility that the remaining sites are cleaved by off-target RGEN activity (even though their sequences are quite different from the on-target sequences), RGENs were incubated with PCR amplicons that contain the potential off-target sites in vitro (Figure 19). No cleavage was detected, suggesting that these sites are not off-target sites. It should be noted, however, that exome sequencing analysis is limited by high false negative results (KARAKOC *et al.* 2012): off-target indels present in clones might have been missed.

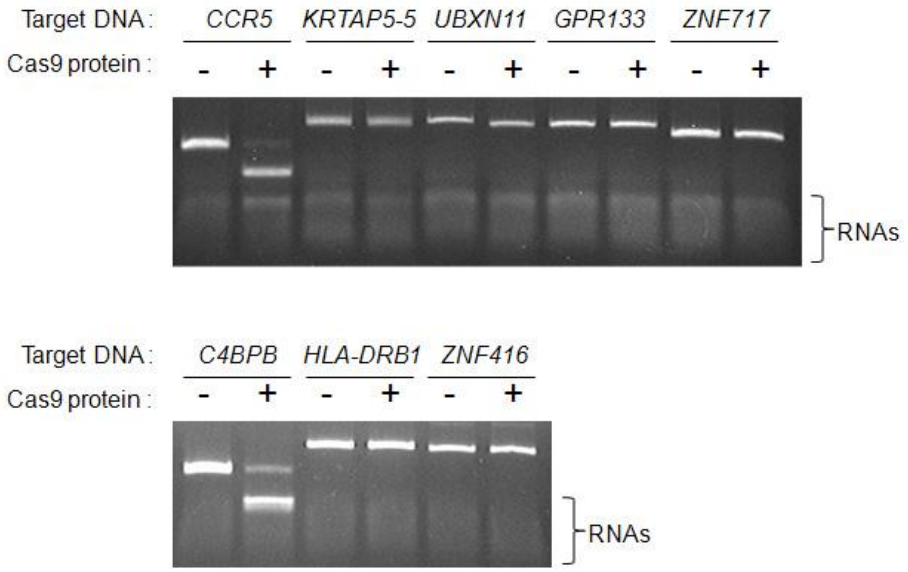


Figure 19. In vitro cleavage assay for potential off-target sequences found in mutant clones. The *CCR5*- (Upper) or *C4BPB*-specific (Bottom) guide-RNA complexed with Cas9 protein was incubated with the PCR product from wild-type cells, which spanned the genomic region of interest. Target sites are listed Table X

Table 7. Sequence alignment to search for putative off-target sites around the indels identified by exome sequencing. Note that most of these putative sites lack the PAM sequence and that indels do not occur at expected positions (3 base upstream of the PAM sequence), suggesting that these indels were not caused by RGEN off-target effects.

Target gene	Indel locus		SEQUENCE ALIGNMENT*	# of matched bases	*p*-value**
CCR5	*CRIPAK*	alignment	CCgATGTATAATAATTGATGTC----A	8	1.0×10^{-1}
		Reference	ACACACGTGCCCATGTGGAGTGCCCGCCTGC----TCACGTGCCCATGTGGAGTGCCCGCCTGCTCACACGTGCCG		
		Altered	ACACACGTGCCCATGTGGAGTGCCCGCCTGCCACATCACGTGCCCATGTGGAGTGCCCGCCTGCTCACACGTGCCG		
	ZNF703	alignment	CCgATGTATAATAATTGATGTCA CCgATGTATAATAATTGATGTCA TGACATCAATTATTATACATcGG	6	3.8×10^{-1}
		Reference	CACCTCGGCGGCTCCAGCTGCTCCACCTGCAGCGCGCACGACCCTGCCGGGCCCAGCCTGAAGGCGGGGGGCTACC		
		Altered	CACCTCGGCGGCTCCAGCTGCTCCACCTGCAGCGCGC-CGACCCTGCCGGGCCCAGCCTGAAGGCGGGGGGCTACC		
	GPR133	alignment	CCgATGTATAATAATTGATGTCA TGACATCAATTATTATACATcGG TGACATCAATTATTATACATcGG	10	1.4×10^{-2}
		Reference	AAAGCTGCTGCGGCTGTGCTGCTGGTACTCCTGGCTGCTGCTATTTTATTACAACTTTCAGGTAATGTCCCCAAGT		
		Altered	AAAGCTGCTGCGGCTGTGCTGCTGG----------TATTTTATTACAACTTTCAGGTAATGTCCCCAAGT		
	TPTE	alignment	TGACATCAATTATTATACATcGG	10	1.4×10^{-2}
		Reference	ATATTTTTGTGAATCATAAGCATGTTTTGCCCACTTACCCGCCTTCTTATCAGCTTTTCAAGTTGTCTTTTTTGAT		
		Altered	ATATTTTTGTGAATCATAAGCATGTTTTGCCCACTTACCCGC---CTTATCAGCTTTTCAAGTTGTCTTTTTTGAT		
	PRAMEF4	alignment	TGACATCAATTATTATAC-ATcGG	9	4.1×10^{-2}
		Reference	AATCTGCTCCAGCAGAGAGTACCATCAGCACCATAACTCTCC-CGGGGGGCAGGATACAGCTCCACGCATAAGTTT		
		Altered	AATCTGCTCCAGCAGAGAGTACCATCAGCACCATAACTCTCCGCGGGGGGCAGGATACAGCTCCACGCATAAGTTT		
	UBXN11	alignment	CCgATGTAT------AATAATTGATGTCA	7	2.1×10^{-1}
		Reference	GACAGGGACTGGGGCCGGGACCGGGAC------CGGGACTGGGGCCGGGACCGGGACTGGGGCCGGGACC		
		Altered	GACAGGGACTGGGGCCGGGACCGGGACGGGACACGGGACTGGGGCCGGGACCGGGACTGGGGCCGGGACC		
	KRTAP5-5	alignment	CCgATGTAT-----------AATAATTGATGTCA	8	1.0×10^{-1}
		Reference	TCTGGCTGTGGGGGCTGTGGCTCCGGCTGTGG-----------AGGCTGTGGGGGCTGTGGCTCCGGCTGTGCGG		
		Altered	TCTGGCTGTGGGGGCTGTGGCTCCGGCTGTGGGGCTGTGGCTCCAGGCTGTGGGGGCTGTGGCTCCGGCTGTGCGG		
	GRIN2C	alignment	CCgATGTATAAT---------AATTGATGTCA TGACATCAATTATTAT---------ACATcGG	7	2.1×10^{-1}
		Reference	CAGCTGCTCCGGACCGAGCAGCGGCAGGTCCTCC---------AGCTCCGGGAAGAGCGGGAGGAAGGGGCGGCCG		
		Altered	CAGCTGCTCCGGACCGAGCAGCGGCAGGTCCTCCGCTCCGGGGAGCTCCGGGAAGAGCGGGAGGAAGGGGCGGCCG		
	ZNF717	alignment	CCgATGTATAATAATTGATGTCA TGACATCAATTATTATACATcGG TGACATCAATTATTATACATcGG TGACATCAATTATTATACATcGG TGACATCAATTATTATACATcGG TGACATCAATTATTATACATcGG TGACATCAATTATTATACATcGG	9	4.1×10^{-2}
		Reference	TGATGTACAAAGAGTTTTGACTTCTGGGAGAAGGTTTTCCTACATTCTTTACATTCAAATGGCTTCTCACCTGTGT		
		Altered	TGATGTACAAAGAGTTTTGACTTCTGGGAGAAGGTTTTC---------TTACATTCAAATGGCTTCTCACCTGTGT		
	MUC4	alignment	CCgATGTATAAT----------AATTGATGTCA	11	3.9×10^{-3}
		Reference	AGGATGCTGAGGAAG----------AGCTGGTGACAGG		
		Altered	AGGATGCTGAGGAAGGCTGGTGACAGGAAGAGGGGTGGTGTGACCTGAGGATGCTGAGGAAGGAGCTGGTGACAGG		
C4BPB	*HLA-DRB1*	alignment	AATGACCACTACAT-CCTCAAgGG CCcTTGAGGATGTAGTGGTCATT	9	4.1×10^{-2}
		Reference	CACGCTTGGGTGCTCCCACTTGGCAGGTGTAAACCTCTCCACTT-CGAGGAACTGTTTCCAGCATCACCAGGGTCTG		
		Altered	CACGCTTGGGTGCTCCCACTTGGCAGGTGTAAACCTCTCCACTTTCGAGGAACTGTTTCCAGCATCACCAGGGTCTG		
	FAM186A	alignment	AATGACCACTACATCCTCAAgGG	11	3.9×10^{-3}
		Reference	CAGGGCCTGGACCTGCTGAGGGGTGAGAGGGATCCCCAATTCCTGAGCCTGCGGAGGGATCAGAGGGATCCCCAGG		
		Altered	CAGGGCCTGGACCTGCTGAGGGGTG----------AGAGGGATCCCCAGG		
	ZNF416	alignment	C-CcTTGAGGATGTAGTGGTCATT CCcTTGAGGATGTAGTGGT-CATT AATGACCACTACATCCT-CAAgGG	10	1.4×10^{-2}
		Reference	CTCATATGGCCTTTCTCCTGTGTGAACCACTTGGTGTT-GAATGAGGCTAGACTTTTGGATAAAGGATTTCCCGCA		
		Altered	CTCATATGGCCTTTCTCCTGTGTGAACCACTTGGTGTTAGAATGAGGCTAGACTTTTGGATAAAGGATTTCCCGCA		

* Indels found in gene-disrupted clones and their corresponding sites in the reference genome are shown. The PAM sequence is shown in blue.

** Note that the cut-off *p*-value is 10^{-5} for the BLAST search. None of these sites are associated with the on-target sites.

IV. Discussion

In this study, human genome was mutated in a targeted manner using Cas9 endonuclease, derived from *Streptococcus pyogenes*. It showed that targeted DNA sequences mutated at a frequency of up to 75%. Although the general success rate of RGEN was not addressed, all 13 RGEN used were could induce mutations in the human genome with various efficiencies.

ZFNs and TALENs enable targeted mutagenesis in a variety of eukaryotic cells and metazoans, but the mutation frequencies obtained with individual nuclease vary widely. Furthermore, some ZFNs and TALENs fail to show any genome editing activities (KIM *et al.* 2010; REYON *et al.* 2012). However, it is technically challenging and time-consuming to make custom nucleases. In this manner, RGENs have a merit compared to ZFNs or TALENs. RGENs can be more readily customized because only the RNA component needs to be replaced to make a new efficient rare-cutting nuclease. Very recently, RGENs were used to manipulate the genome in various cells (CHO *et al.* 2013a; CONG *et al.* 2013; DICARLO *et al.* 2013; DING *et al.* 2013; JIANG *et al.* 2013; JINEK *et al.* 2013; MALI *et al.* 2013b; REYON *et al.* 2012; SUNG *et al.* 2013) and organisms (CHANG *et al.* 2013; CHO *et al.* 2013c; FRIEDLAND *et al.* 2013; HWANG *et al.* 2013; LI *et al.* 2013a; LI *et al.* 2013b; LI *et al.* 2013c; NEKRASOV *et al.* 2013; SEBO *et al.* 2013; SHAN *et al.* 2013; WANG *et al.*

2013). Furthermore, with the development of technology for synthesizing long ribonucleotides, RGEN is a eligible candidate for customizable, ready-to-use endonuclease.

Applications of engineered nucleases in research, biotechnology, and medicine are hampered by their off-target effects. Both ZFNs and TALENs can induce mutations at off-target sites that are highly homologous to on-target sites (KIM *et al.* 2012; MUSSOLINO *et al.* 2011). Too much off-target DNA cleavage causes cellular toxicity. Furthermore, the repair of off-target DSBs gives rise to unwanted chromosomal aberrations such as deletions, inversions, and translocations (LEE *et al.* 2010; LEE *et al.* 2012).

Unlike ZFNS and TALENs, whose specificities of DNA are determined by protein-DNA interactions that are often DNA context-dependent (SANDER *et al.* 2011) and not easy to predict, RGENs recognize target sites mostly by the simple Watson-Crick base-pairing rules. For example, ZFNs induce off-target mutations at cryptic sites that cannot be predicted based on sequence homology alone (GABRIEL *et al.* 2011). However, RGENs possibly do not cleave off-target DNA sequences whose recognition does not follow the base-pairing rule. Thus, it is important to address the specificity of RGENs that how many mismatches RGENs can tolerate to induce mutations at off-target sites (CHO *et al.* 2013b; FU *et al.* 2013; HSU *et al.* 2013; PATTANAYAK *et al.* 2013).

As determined in this study, RGENs induce off-target mutations at sites with a single-base mismatch. However, RGENs efficiently discriminate on-

target sites from off-target sites that differ by only two or three bases. Furthermore, exome sequencing showed that no off-target mutations were present in four clonal populations of mutant cells. These results suggest that choice of a unique target site will ensure the precise and safe genome manipulation at a desired locus. In principle, specificity of RGENs are fixed in length and are greater than human genome ($4^{20} \times 4^2 = 1.8 \times 10^{13} > 3.0 \times 10^9$). It is still sufficient to ensure targeting of most genes of human genome if considering homologous sites with two base mismatches. For example, there are 28 targetable sequences in *CCR5* which have no homologous sites including two or less mismatches (Table 8). It is possible that the 11 RGENs induce off-target mutations at sites not examined here, which could be revealed by deep sequencing at other less-homologous candidate sites or by whole genome sequencing. In addition, in vitro selection of cleavage sites (PATTANAYAK *et al.* 2011) or the IDLV capture approach (GABRIEL *et al.* 2011) used for the identification of ZFN off-target sites may reveal cryptic sites cleaved by RGENs.

Although PAM sequence of Cas9 derived from *S.pyogenes* essentially appear every eight base pairs ($4 \times 4 \times 2^{-1}$) in genome, it possibly limits the choosing target sequences. This limitation may be overcome by using other Cas9 variants (ESVELT *et al.* 2013; HOU *et al.* 2013) or engineered Cas9.

Cas9 can be used for a diverse application in molecular biology as well as ZFNs or TALENs. Nicking Cas9 derived from inactivation one of the two

active residue, will promise for more safe genome manipulation (MALI *et al.* 2013a). Using two nicking Cas9 in pair will double the specificity of Cas9-mediated mutagenesis (CHO *et al.* 2013b; RAN *et al.* 2013). Indeed, inactivated Cas9 can be used as a scaffold for delivering other functional domains such as transcriptional activator or repressor (GILBERT *et al.* 2013; MAEDER *et al.* 2013; PEREZ-PINERA *et al.* 2013; QI *et al.* 2013), or methyltransferase (MEISTER *et al.* 2010) to chromosome.

In summary, prokaryotic Cas9 endonuclease can function in human cells with high efficiency and specificity. Unlike ZFNs or TALENs, the specificity of RGENs can be customized by replacing a synthetic RNA molecule without changing the protein component. Thus, RGEN technology will be a powerful option that will facilitate precision genome editing in biotechnology, research and medicine.

Table 8. List of unique targetable sequences in *CCR5*.

Target Seq	Position	Strand	mis-0	mis-1	mis-2	mis-3
TCATCCTGATAAACTGCAAAAGG	158	+	1	0	0	24
TCACTATGCTGCCGCCCAGTGGG	261	+	1	0	0	5
ACAATGTGTCAACTCTTGACAGG	295	+	1	0	0	3
TTGACAGGGCTCTATTTTATAGG	310	+	1	0	0	6
TATTTTATAGGCTTCTTCTCTGG	322	+	1	0	0	22
TCATCCTCCTGACAATCGATAGG	356	+	1	0	0	2
GGTGACAAGTGTGATCACTTGGG	438	+	1	0	0	12
TTTACCAGATCTCAAAAAGAAGG	496	+	1	0	0	14
CATTAAAGATAGTCATCTTGGGG	584	+	1	0	0	19
AAAGATAGTCATCTTGGGGCTGG	588	+	1	0	0	15
ATAATTGCAGTAGCTCTAACAGG	800	+	1	0	0	10
CAGGTTGGACCAAGCTATGCAGG	819	+	1	0	0	5
CTGTTCTATTTTCCAGCAAGAGG	969	+	1	0	0	13
TCAGTTTACACCCGATCCACTGG	1009	+	1	0	0	1
CAGTTTACACCCGATCCACTGGG	1010	+	1	0	0	2
AGTTTACACCCGATCCACTGGGG	1011	+	1	0	0	2
CCTGCCAAAAAATCAATGTGAAG	56	-	1	0	0	29
CCTCCTGCCTCCGCTCTACTCAC	93	-	1	0	0	40
CCTGATAAACTGCAAAAGGCTGA	162	-	1	0	0	15
CCTTCTGGGCTCACTATGCTGCC	251	-	1	0	0	24
CCCAGTGGGACTTTGGAAATACA	275	-	1	0	0	15
CCAGGAATCATCTTTACCAGATC	484	-	1	0	0	18
CCCTACAACATTGTCCTTCTCCT	748	-	1	0	0	22
CCTACAACATTGTCCTTCTCCTG	749	-	1	0	0	25
CCCCATCATCTATGCCTTTGTCG	879	-	1	0	0	10
CCAAAAGCACATTGCCAAACGCT	936	-	1	0	0	6
CCAAACGCTTCTGCAAATGCTGT	950	-	1	0	0	13
CCACTGGGGAGCAGGAAATATCT	1025	-	1	0	0	29

VI. References

BAE, K. H., Y. D. KWON, H. C. SHIN, M. S. HWANG, E. H. RYU *et al.*, 2003 Human zinc fingers as building blocks in the construction of artificial transcription factors. Nat Biotechnol **21:** 275-280.

BIBIKOVA, M., K. BEUMER, J. K. TRAUTMAN and D. CARROLL, 2003 Enhancing gene targeting with designed zinc finger nucleases. Science **300:** 764.

CAPECCHI, M. R., 1989 Altering the genome by homologous recombination. Science **244:** 1288-1292.

CHANG, N., C. SUN, L. GAO, D. ZHU, X. XU *et al.*, 2013 Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. Cell Res **23:** 465-472.

CHO, S. W., S. KIM, J. M. KIM and J. S. KIM, 2013a Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol **31:** 230-232.

CHO, S. W., S. KIM, Y. KIM, J. KWEON, H. S. KIM *et al.*, 2013b Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res.

CHO, S. W., J. LEE, D. CARROLL and J. S. KIM, 2013c Heritable Gene Knockout in Caenorhabditis elegans by Direct Injection of Cas9-

sgRNA Ribonucleoproteins. Genetics **195:** 1177-1180.

CLARK, M. J., R. CHEN, H. Y. LAM, K. J. KARCZEWSKI, G. EUSKIRCHEN *et al.*, 2011 Performance comparison of exome DNA sequencing technologies. Nat Biotechnol **29:** 908-914.

CONG, L., F. A. RAN, D. COX, S. LIN, R. BARRETTO *et al.*, 2013 Multiplex genome engineering using CRISPR/Cas systems. Science **339:** 819-823.

CUI, X., D. JI, D. A. FISHER, Y. WU, D. M. BRINER *et al.*, 2011 Targeted integration in rat and mouse embryos with zinc-finger nucleases. Nat Biotechnol **29:** 64-67.

DELTCHEVA, E., K. CHYLINSKI, C. M. SHARMA, K. GONZALES, Y. CHAO *et al.*, 2011 CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature **471:** 602-607.

DICARLO, J. E., J. E. NORVILLE, P. MALI, X. RIOS, J. AACH *et al.*, 2013 Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic Acids Res **41:** 4336-4343.

DING, Q., S. N. REGAN, Y. XIA, L. A. OOSTROM, C. A. COWAN *et al.*, 2013 Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. Cell Stem Cell **12:** 393-394.

DOYON, Y., J. M. MCCAMMON, J. C. MILLER, F. FARAJI, C. NGO *et al.*, 2008 Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. Nat Biotechnol **26:** 702-708.

ESVELT, K. M., P. MALI, J. L. BRAFF, M. MOOSBURNER, S. J. YAUNG *et al.*, 2013 Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. Nat Methods **10:** 1116-1121.

FRIEDLAND, A. E., Y. B. TZUR, K. M. ESVELT, M. P. COLAIACOVO, G. M. CHURCH *et al.*, 2013 Heritable genome editing in C. elegans via a CRISPR-Cas9 system. Nat Methods.

FU, Y., J. A. FODEN, C. KHAYTER, M. L. MAEDER, D. REYON *et al.*, 2013 High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol **31:** 822-826.

GABRIEL, R., A. LOMBARDO, A. ARENS, J. C. MILLER, P. GENOVESE *et al.*, 2011 An unbiased genome-wide analysis of zinc-finger nuclease specificity. Nat Biotechnol **29:** 816-823.

GEURTS, A. M., G. J. COST, Y. FREYVERT, B. ZEITLER, J. C. MILLER *et al.*, 2009 Knockout rats via embryo microinjection of zinc-finger nucleases. Science **325:** 433.

GEURTS, A. M., Y. YANG, K. J. CLARK, G. LIU, Z. CUI *et al.*, 2003 Gene transfer into genomes of human cells by the sleeping beauty transposon system. Mol Ther **8:** 108-117.

GILBERT, L. A., M. H. LARSON, L. MORSUT, Z. LIU, G. A. BRAR *et al.*, 2013 CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell **154:** 442-451.

HACEIN-BEY-ABINA, S., C. VON KALLE, M. SCHMIDT, M. P. MCCORMACK, N. WULFFRAAT *et al.*, 2003 LMO2-associated clonal T cell

proliferation in two patients after gene therapy for SCID-X1. Science **302:** 415-419.

HOCKEMEYER, D., F. SOLDNER, C. BEARD, Q. GAO, M. MITALIPOVA *et al.*, 2009 Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nat Biotechnol **27:** 851-857.

HOU, Z., Y. ZHANG, N. E. PROPSON, S. E. HOWDEN, L. F. CHU *et al.*, 2013 Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. Proc Natl Acad Sci U S A **110:** 15644-15649.

HSU, P. D., D. A. SCOTT, J. A. WEINSTEIN, F. A. RAN, S. KONERMANN *et al.*, 2013 DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol **31:** 827-832.

HWANG, W. Y., Y. FU, D. REYON, M. L. MAEDER, S. Q. TSAI *et al.*, 2013 Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol **31:** 227-229.

JASIN, M., 1996 Genetic manipulation of genomes with rare-cutting endonucleases. Trends Genet **12:** 224-228.

JIANG, W., D. BIKARD, D. COX, F. ZHANG and L. A. MARRAFFINI, 2013 RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol **31:** 233-239.

JINEK, M., K. CHYLINSKI, I. FONFARA, M. HAUER, J. A. DOUDNA *et al.*, 2012 A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science **337:** 816-821.

JINEK, M., A. EAST, A. CHENG, S. LIN, E. MA *et al.*, 2013 RNA-programmed genome editing in human cells. Elife **2:** e00471.

KARAKOC, E., C. ALKAN, B. J. O'ROAK, M. Y. DENNIS, L. VIVES *et al.*, 2012 Detection of structural variants and indels within exome data. Nat Methods **9:** 176-178.

KIM, E., S. KIM, D. H. KIM, B. S. CHOI, I. Y. CHOI *et al.*, 2012 Precision genome engineering with programmable DNA-nicking enzymes. Genome Res **22:** 1327-1333.

KIM, H. J., H. J. LEE, H. KIM, S. W. CHO and J. S. KIM, 2009 Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. Genome Res **19:** 1279-1288.

KIM, J. S., H. J. LEE and D. CARROLL, 2010 Genome editing with modularly assembled zinc-finger nucleases. Nat Methods **7:** 91; author reply 91-92.

KIM, J. S., and C. O. PABO, 1998 Getting a handhold on DNA: design of poly-zinc finger proteins with femtomolar dissociation constants. Proc Natl Acad Sci U S A **95:** 2812-2817.

KIM, Y., J. KWEON, A. KIM, J. K. CHON, J. Y. YOO *et al.*, 2013 A library of TAL effector nucleases spanning the human genome. Nat Biotechnol **31:** 251-258.

KIM, Y. G., J. CHA and S. CHANDRASEGARAN, 1996 Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci U S A **93:** 1156-1160.

LEE, H. J., E. KIM and J. S. KIM, 2010 Targeted chromosomal deletions in human cells using zinc finger nucleases. Genome Res **20:** 81-89.

LEE, H. J., J. KWEON, E. KIM, S. KIM and J. S. KIM, 2012 Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. Genome Res **22:** 539-548.

LI, D., Z. QIU, Y. SHAO, Y. CHEN, Y. GUAN *et al.*, 2013a Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. Nat Biotechnol **31:** 681-683.

LI, J. F., J. E. NORVILLE, J. AACH, M. MCCORMACK, D. ZHANG *et al.*, 2013b Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. Nat Biotechnol **31:** 688-691.

LI, W., F. TENG, T. LI and Q. ZHOU, 2013c Simultaneous generation and germline transmission of multiple gene mutations in rat using CRISPR-Cas systems. Nat Biotechnol **31:** 684-686.

LOBO, S. M., S. IFILL and N. HERNANDEZ, 1990 cis-acting elements required for RNA polymerase II and III transcription in the human U2 and U6 snRNA promoters. Nucleic Acids Res **18:** 2891-2899.

MAEDER, M. L., S. J. LINDER, V. M. CASCIO, Y. FU, Q. H. HO *et al.*, 2013 CRISPR RNA-guided activation of endogenous human genes. Nat Methods **10:** 977-979.

MAEDER, M. L., S. THIBODEAU-BEGANNY, A. OSIAK, D. A. WRIGHT, R. M. ANTHONY *et al.*, 2008 Rapid "open-source" engineering of

customized zinc-finger nucleases for highly efficient gene modification. Mol Cell **31:** 294-301.

MALI, P., J. AACH, P. B. STRANGES, K. M. ESVELT, M. MOOSBURNER *et al.*, 2013a CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol **31:** 833-838.

MALI, P., L. YANG, K. M. ESVELT, J. AACH, M. GUELL *et al.*, 2013b RNA-guided human genome engineering via Cas9. Science **339:** 823-826.

MEISTER, G. E., S. CHANDRASEGARAN and M. OSTERMEIER, 2010 Heterodimeric DNA methyltransferases as a platform for creating designer zinc finger methyltransferases for targeted DNA methylation in cells. Nucleic Acids Res **38:** 1749-1759.

MENG, X., M. B. NOYES, L. J. ZHU, N. D. LAWSON and S. A. WOLFE, 2008 Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. Nat Biotechnol **26:** 695-701.

MILLER, A. D., and G. J. ROSMAN, 1989 Improved retroviral vectors for gene transfer and expression. Biotechniques **7:** 980-982, 984-986, 989-990.

MILLER, J. C., M. C. HOLMES, J. WANG, D. Y. GUSCHIN, Y. L. LEE *et al.*, 2007 An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotechnol **25:** 778-785.

MILLER, J. C., S. TAN, G. QIAO, K. A. BARLOW, J. WANG *et al.*, 2011 A TALE nuclease architecture for efficient genome editing. Nat Biotechnol **29:** 143-148.

MOJICA, F. J., C. DIEZ-VILLASENOR, J. GARCIA-MARTINEZ and C. ALMENDROS, 2009 Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology **155:** 733-740.

MORTON, J., M. W. DAVIS, E. M. JORGENSEN and D. CARROLL, 2006 Induction and repair of zinc-finger nuclease-targeted double-strand breaks in Caenorhabditis elegans somatic cells. Proc Natl Acad Sci U S A **103:** 16370-16375.

MUSSOLINO, C., R. MORBITZER, F. LUTGE, N. DANNEMANN, T. LAHAYE *et al.*, 2011 A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Res **39:** 9283-9293.

NEKRASOV, V., B. STASKAWICZ, D. WEIGEL, J. D. JONES and S. KAMOUN, 2013 Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. Nat Biotechnol **31:** 691-693.

O'DRISCOLL, M., and P. A. JEGGO, 2006 The role of double-strand break repair - insights from human genetics. Nat Rev Genet **7:** 45-54.

OUSTEROUT, D. G., P. PEREZ-PINERA, P. I. THAKORE, A. M. KABADI, M. T. BROWN *et al.*, 2013 Reading frame correction by targeted genome editing restores dystrophin expression in cells from Duchenne muscular dystrophy patients. Mol Ther **21:** 1718-1726.

PATTANAYAK, V., S. LIN, J. P. GUILINGER, E. MA, J. A. DOUDNA *et al.*, 2013 High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat Biotechnol **31:** 839-843.

PATTANAYAK, V., C. L. RAMIREZ, J. K. JOUNG and D. R. LIU, 2011 Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. Nat Methods **8:** 765-770.

PEREZ-PINERA, P., D. D. KOCAK, C. M. VOCKLEY, A. F. ADLER, A. M. KABADI *et al.*, 2013 RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat Methods **10:** 973-976.

PEREZ, E. E., J. WANG, J. C. MILLER, Y. JOUVENOT, K. A. KIM *et al.*, 2008 Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nat Biotechnol **26:** 808-816.

PRUETT-MILLER, S. M., J. P. CONNELLY, M. L. MAEDER, J. K. JOUNG and M. H. PORTEUS, 2008 Comparison of zinc finger nucleases for use in gene targeting in mammalian cells. Mol Ther **16:** 707-717.

QI, L. S., M. H. LARSON, L. A. GILBERT, J. A. DOUDNA, J. S. WEISSMAN *et al.*, 2013 Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell **152:** 1173-1183.

RAN, F. A., P. D. HSU, C. Y. LIN, J. S. GOOTENBERG, S. KONERMANN *et al.*, 2013 Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell **154:** 1380-1389.

REBAR, E. J., and C. O. PABO, 1994 Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. Science **263:** 671-673.

REYON, D., S. Q. TSAI, C. KHAYTER, J. A. FODEN, J. D. SANDER *et al.*, 2012 FLASH assembly of TALENs for high-throughput genome editing. Nat Biotechnol **30:** 460-465.

ROUET, P., F. SMIH and M. JASIN, 1994 Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. Proc Natl Acad Sci U S A **91:** 6064-6068.

ROUILLARD, J. M., W. LEE, G. TRUAN, X. GAO, X. ZHOU *et al.*, 2004 Gene2Oligo: oligonucleotide design for in vitro gene synthesis. Nucleic Acids Res **32:** W176-180.

RUBIN, G. M., and A. C. SPRADLING, 1982 Genetic transformation of Drosophila with transposable element vectors. Science **218:** 348-353.

SANDER, J. D., E. J. DAHLBORG, M. J. GOODWIN, L. CADE, F. ZHANG *et al.*, 2011 Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). Nat Methods **8:** 67-69.

SANJANA, N. E., L. CONG, Y. ZHOU, M. M. CUNNIFF, G. FENG *et al.*, 2012 A transcription activator-like effector toolbox for genome engineering. Nat Protoc **7:** 171-192.

SCHRODER, A. R., P. SHINN, H. CHEN, C. BERRY, J. R. ECKER *et al.*, 2002 HIV-1 integration in the human genome favors active genes and local hotspots. Cell **110:** 521-529.

SEBO, Z. L., H. B. LEE, Y. PENG and Y. GUO, 2013 A simplified and efficient germline-specific CRISPR/Cas9 system for Drosophila genomic engineering. Fly (Austin) **8**.

SEDIVY, J. M., and P. A. SHARP, 1989 Positive genetic selection for gene disruption in mammalian cells by homologous recombination. Proc Natl Acad Sci U S A **86:** 227-231.

SHAN, Q., Y. WANG, J. LI, Y. ZHANG, K. CHEN *et al.*, 2013 Targeted genome modification of crop plants using a CRISPR-Cas system. Nat Biotechnol **31:** 686-688.

SHUKLA, V. K., Y. DOYON, J. C. MILLER, R. C. DEKELVER, E. A. MOEHLE *et al.*, 2009 Precise genome modification in the crop species Zea mays using zinc-finger nucleases. Nature **459:** 437-441.

SUNG, Y. H., J. M. KIM, H. T. KIM, J. LEE, J. JEON *et al.*, 2013 Highly efficient gene knockout in mice and zebrafish with RNA-guided endonucleases. Genome Res.

TOWNSEND, J. A., D. A. WRIGHT, R. J. WINFREY, F. FU, M. L. MAEDER *et al.*, 2009 High-frequency modification of plant genes using engineered zinc-finger nucleases. Nature **459:** 442-445.

URNOV, F. D., J. C. MILLER, Y. L. LEE, C. M. BEAUSEJOUR, J. M. ROCK *et al.*, 2005 Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature **435:** 646-651.

URNOV, F. D., E. J. REBAR, M. C. HOLMES, H. S. ZHANG and P. D. GREGORY, 2010 Genome editing with engineered zinc finger nucleases. Nat Rev Genet **11:** 636-646.

WANG, H., H. YANG, C. S. SHIVALILA, M. M. DAWLATY, A. W. CHENG *et al.*, 2013 One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell **153:** 910-918.

WIEDENHEFT, B., S. H. STERNBERG and J. A. DOUDNA, 2012 RNA-guided

genetic silencing systems in bacteria and archaea. Nature **482:** 331-338.

WU, X., Y. LI, B. CRISE and S. M. BURGESS, 2003 Transcription start regions in the human genome are favored targets for MLV integration. Science **300:** 1749-1751.

YANG, D., H. YANG, W. LI, B. ZHAO, Z. OUYANG *et al.*, 2011 Generation of PPARgamma mono-allelic knockout pigs via zinc-finger nucleases and nuclear transfer cloning. Cell Res **21:** 979-982.

YU, S., J. LUO, Z. SONG, F. DING, Y. DAI *et al.*, 2011 Highly efficient modification of beta-lactoglobulin (BLG) gene via zinc-finger nucleases in cattle. Cell Res **21:** 1638-1640.

국문 초록

세균이나 고세균에서 발견되는 Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated (CRISPR/Cas) 시스템은 박테리오파지나 플라스미드와 같은 외부에서 도입된 유전 물질을 절단해서 면역 작용을 하는 것으로 알려져 있다. CRISPR/Cas type II 시스템은 3 가지로 구성된다. 효소로 작용하는 Cas9 단백질, DNA 와 결합하는 crRNA, 그리고 이 tracrRNA 가 존재한다. Cas9 단백질은 crRNA, tracrRNA 와 함께 복합체를 이루어 제한 효소의 활성을 나타내는 것으로 알려져 있다. 기존의 연구를 통해, crRNA 염기 서열 중 일부가 20 여 개의 DNA 염기 서열을 인식해 상보적으로 결합 할 수 있으며, 이 부분의 염기서열을 조작하면 원하는 서열을 가지는 DNA 를 자를 수 있도록 만들 수 있다는 것이 알려져 있다.

본 연구에서는 CRISPR/Cas 를 이용해 인간 배양 세포에서 선택적으로 돌연변이를 일으키는데 성공하였으며, 다양한 염기 서열에 대해서 검증하였다. 뿐만 아니라 인간 유전체에서 유사한 염기 서열에서 비특이적인 돌연변이가 일어 나지 않는 것을 확인 하였다. RGEN 의 경우 손 쉽게 많은 유전자 가위를 만들 수 있고 적용할 수 있기 때문에 과학 연구, 바이오 산업 및 치료제 개발과 같은 의학 분야에 크게 기여 할 것이다.

학 번 : 2007-20326